DNA Confined in Nanochannels and Nanoslits

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Douglas R. Tree

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Kevin D. Dorfman

May 2014

Acknowledgements

At this bittersweet close to my doctoral studies, I would like to thank my advisor, Kevin Dorfman for his support. He has truly given me every academic opportunity that I could have imagined five years ago. Indeed, he has not only provided generous financial support and a stimulating academic environment, but he has also given me a tremendous amount of freedom to persue my nascent academic interests. Moreover, he has been a counselor and a friend, and our many discussions have challenged me to think deeper, work harder and aspire to loftier goals. Thank you Kevin.

In addition, I would like to acknowledge current and former members of the Dorfman group for their friendship and for countless thought-provoking conversations: Damini Gupta, Aashish Jain, Scott King, Nabil Laachi, Maggie Linak, Dan Olson, Jia Ou, Teresa Sells, Julian Sheats, and Joel Thomas. I would also like to specially thank group members Yanwei Wang, Wes Reinhart, Abhiram Muralidhar and Sarit Dutta. Yanwei helped me learn coarse-grained simulations and polymer physics at the beginning of my tenure as a graduate student, and Wes and I worked closely together on a number of projects. Abhiram spent many hours with me discussing research ideas surrounding PERM, and Sarit was an invaluable resource for navigating the world of high performance computing.

I would certainly be remiss if I didn't acknowledge the support of my family, chiefly my wife Tedi. In addition to caring for our growing family of now three children, she has patiently endured tight financial constraints, long business trips and unexecused dinner absences. She has been a source of emotional support during anxious hours in preparation for exams and presentations, and I would not have been able to accomplish this task without her. Thank you Tedi, I love you. I would also like to acknowledge my parents, who taught me the joy of learning at an early age, and who continue to this day to provide wisdom and support for me and my family.

Finally, I would like to acknowledge financial support from the University of Minnesota's Doctoral Dissertation Fellowship, the National Science Foundation (CBET-1262286) and the National Institutes of Health (R01-HG005216, R01-HG006851) as well as computational resources from the Minnesota Supercomputing Institute.

Abstract

It has become increasingly apparent in recent years that next-generation sequencing (NGS) has a blind spot for large scale genomic variation, which is crucial for understanding the genotype-phenotype relationship. Genomic mapping methods attempt to overcome the weakesses of NGS by providing a coarse-grained map of the distances between restriction sites to aid in sequence assembly. From such methods, one hopes to realize fast and inexpensive *de novo* sequencing of human and plant genomes.

One of the most promising methods for genomic mapping involves placing DNA inside a device only a few dozen nanometers wide called a nanochannel. A nanochannel stretches the DNA so that the distance between fluorescently labeled restriction sites can be measured en route to obtaining an accurate genome map. Unfortunately for those who wish to design devices, the physics of how DNA stretches when confined in a nanochannel is still an active area of research. Indeed, despite decades old theories from polymer physics regarding weakly and strongly stretched polymers, seminal experiments in the mid-2000s have gone unexplained until very recently.

With a goal of creating a realistic engineering model of DNA in nanochannels, this dissertation addresses a number of important outstanding research topics in this area. We first discuss the physics of dilute solutions of DNA in free solution, which show distinctive behavior due to the stiff nature of the polymer. We then turn our attention to the equilibrium regimes of confined DNA and explore the effects of stiff chains and weak excluded volume on the confinement free energy and polymer extension. We also examine dynamic properties such as the diffusion coefficient and the characteristic relaxation time. Finally, we discuss a sister problem related to DNA confined in nanoslits, which shares much of the same physics as DNA confined in channels.

Having done this, we find ourselves with a well-parameterized wormlike chain model that is remarkably accurate in describing the behavior of DNA in confinement. As such, it appears that researchers may proceed with the rational design of nanochannel mapping devices using this model.

Contents

List of Tables

List of Figures

Statement of the Author's Contributions

Much of the research contained in this dissertation was performed collaboratively and has been published as co-authored articles in peer-reviewed journals. As such, I would like to clarify my particular contributions.

Portions of Chapters 1 and 2 are based on Sections 2.1 and 7.1-7.3 "Beyond Gel Electrophoresis: Microfluidic Separations, Fluorescence Burst Analysis, and DNA Stretching" *Chem. Rev.* 113 (2013), 8369–8382 by Kevin D. Dorfman, Scott B. King, Daniel W. Olson, Joel D. P. Thomas and Douglas R. Tree. I wrote these sections of the review article.

Portions of Chapter 3 use parts of the supplemental materials to "Is DNA a Good Model Polymer?" *Macromolecules* 46 (2013), 8369–8382 by Douglas R. Tree, Abhiram Muralidhar, Patrick S. Doyle and Kevin D. Dorfman and parts of the supplemental materials to "Extension of DNA in a Nanochannel as a Rod-to-Coil Transition" *Phys. Rev. Lett.* 110 (2013), 208103 by Douglas R. Tree, Yanwei Wang and Kevin D. Dorfman. I wrote both of these supplemental materials.

Chapter 4 is based on "Is DNA a Good Model Polymer?" *Macromolecules* 46 (2013), 8369–8382 by Douglas R. Tree, Abhiram Muralidhar, Patrick S. Doyle and Kevin D. Dorfman. In this paper, we used Pruned-enriched Rosenbluth method (PERM) simulations to look at equilibrium and diffusion properties of free-solution wormlike chains. I wrote the PERM code, performed simulations and analyzed most of the data. Calculations of the structure factor and diffusivity were performed by A. Muralidhar.

Chapter 5 is based on "Extension of DNA in a Nanochannel as a Rod-to-Coil Transition" *Phys. Rev. Lett.* 110 (2013), 208103 by Douglas R. Tree, Yanwei Wang and Kevin D. Dorfman. This article examined the free energy of chains confined in square channels resulting from PERM simulations. I wrote the relevant code, performed simulations and analyzed all of the data. Interpretation of the results was shared jointly by all of the authors.

Chapter 6 is based on "Mobility of a Semiflexible Chain Confined in a Nanochannel" *Phys. Rev. Lett.* 108 (2012), 228105 by Douglas R. Tree, Yanwei Wang and Kevin D.

Dorfman. This article uses Metropolis Monte Carlo (MMC) simulations combined with a comptuational fluid dyanmics (CFD) method for estimating the Green's function of a point source in a square channel to calculate the diffusion coefficient of a confined polymer chain. I wrote the CFD code, combined it with a Monte Carlo code previously written by Y. Wang, and performed the simulations and data analysis. Interpretation of the results was shared jointly by all of the authors.

Chapter 7 is based on "Modeling the Relaxation Time of DNA Confined in a Nanochannel" *Biomicrofluidics* 7 (2013), 054118 by Douglas R. Tree, Yanwei Wang and Kevin D. Dorfman. This article uses the MMC and CFD codes mentioned previously to estimate the relaxation time of a semiflexible chain in a square channel. The inital idea for this project originated with discussions and the dumbbell model derivation by Y. Wang. Subsequently, I performed the simulations and analyzed the data. Interpretation of the results was shared jointly by all of the authors.

Chapter 8 is based on "The Odijk Regime in Slits" currently in review at *Macromolecules* by Douglas R. Tree, Wesley F. Reinhart and Kevin D. Dorfman. This article uses PERM simulations to obtain equilibrium properties of semiflexible chains confined to slits much smaller than the chain persistence length. Much of this work originated with joint efforts between myself and W. Reinhart, who performed preliminary simulations and data analysis. I collected the final data that appears in the paper and contributed to the analysis, specifically the review and formulation of the scaling theory contained therein. Interpretation of the results was shared jointly by all of the authors.

Chapter 1

Introduction

1.1 Next Generation Sequencing

The so-called central dogma of molecular biology states that the information encoded in the nucleotide sequence of the deoxyribonucleic acid (DNA) molecules of an organism (its genome) dictates the transcription of proteins. In turn, proteins determine the phenotype, or set of observable characteristics, expressed by the organism. With an eye towards understanding and controlling phenotypic expression, both students and practitioners of biology and medicine have an extraordinary interest in analyzing and manipulating DNA. Spectacular developments in modern molecular biology have led to many such tools, including the ability to read the sequence of the nucleotides of the entire genome of an organism [1, 2].

Even with the knowledge that molecular biology is not as simple as the central dogma implies [3], there is much hope in the scientific community that studying DNA sequences will lead to an increased understanding of phenotypic expression [4]. Indeed many have speculated that the study of the human genome will lead to deeper insights into genetic diseases [4], an increased understanding of cancer [5] and autism [6], and the possibility of more personalized medical care [7]. Beyond human health, it is hoped that genomic sequencing will also play an important role in improving agricultural products such as corn, rice and wheat [8–10].

Although modern sequencing methods differ in many of their specific details, they share common unit operations, which are outlined in Figure 1.1 [11–13]. In the most

common shotgun-type process, (a) the genome is extracted and (b) randomly sheared into much smaller fragments. These fragments are then (c) multiplied using PCR amplification to increase signal and (d) a length of sequence between 25 and 1000 pairs of nucleotides (or base pairs, bp) is "read" (again the details vary widely). Note that this length is small compared to the size of the genome, which is about 3×10^9 bp for humans [1]. Once sequence reads are obtained, they are (e) assembled into groups called contigs and finished to form a consensus sequence [14, 15].

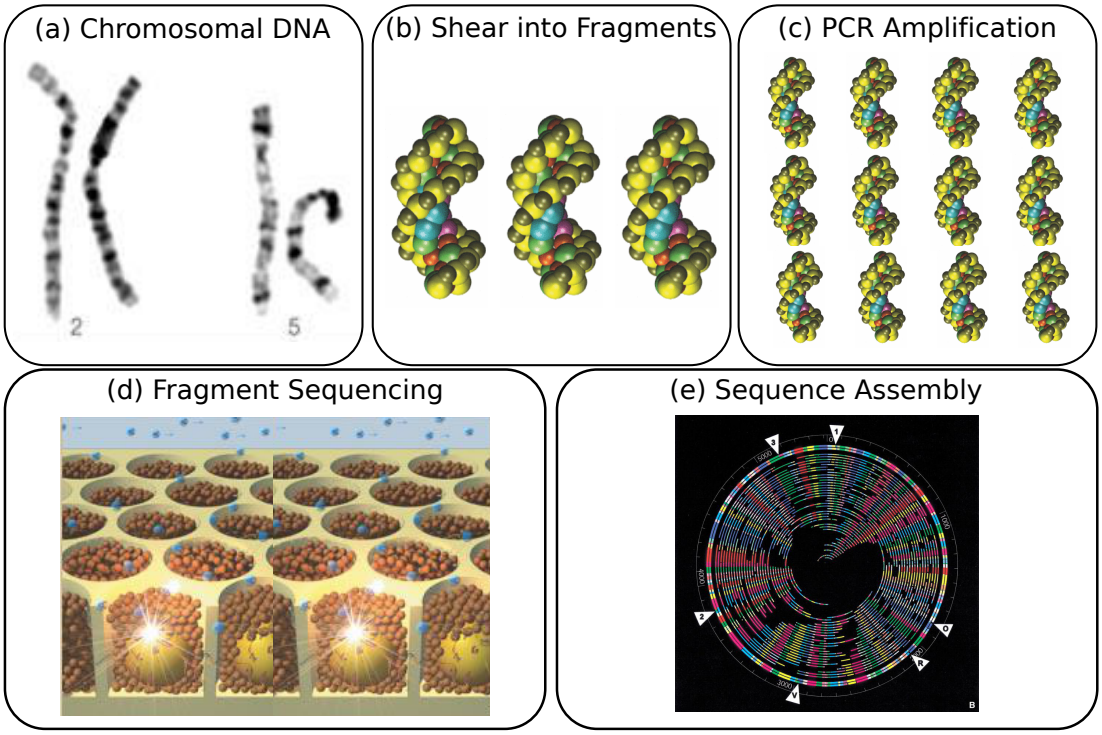

Figure 1.1: Schematic describing the unit operations in next generation sequencing. Images for the schematic have been taken from References 16–19.

In the first generation Sanger sequencing method, the fragments of DNA had to be physically separated by electrophoresis, and automation and parallelization was difficult and expensive [11]. However, innovative "next-generation" sequencing (NGS) processes were designed to read sequences with the goal of automation and massive parallelization specifically in mind [12, 13]. As a result, the cost of sequencing has fallen dramatically and the National Human Genome Research Institute reports a astounding six order of

magnitude decrease in the cost of sequencing a megabase (Mbp) of DNA in just 10 years [20].

Despite this amazing achievement, the ultimate goal of the field is not to obtain a collection of consensus sequences, but rather to catalogue *genetic variation* and draw connections to phenotype [21]. This goal should be juxtaposed with the reality that NGS measurements come from ensembles of molecules originating from copies of chromosomes from many different cells. In fact, a finished consensus sequence is typically comprised of genomes from several different individuals within a species. To study genetic variation however, the gold standard is a *de novo* construction of the sequence unique to each organism or even for a given subset of cells within an organism (*e.g.* cancer cells).

The desire to construct *de novo* sequences highlights a particular weakness of NGS: short read-lengths. In fact, the read-lengths achievable by NGS ($\approx$ 500bp) [12, 13] are even shorter than those possible by Sanger sequencing ($\approx$ 1000 bp) [11], which is already orders of magnitude shorter than the length of the genome. Short read-lengths pose an inherent challenge in sequencing, since the read-length sets the fundamental scale of resolution of the sequencing method. Indeed, a short read-length represents a greater loss of information from the original genome, since two short reads does not equal the information quality of a read twice as long. As a consequence, it is practically impossible for an ensemble-based method with read-lengths less than 1 kilobase pair (kbp) to asses repeat regions and genomic variability on sequence scales much longer than 1 kbp. Or as one researcher bluntly summarized it, "procedures for mapping and sequencing DNA were originally designed to analyze genes rather than genomes [22]."

However, large-scale genomic variation is proving to be particularly relevant in the quest to decode the genotype–phenotype relationship. Recent research suggests that inversions, rearrangements, insertions and deletions of sequences of more than 1 kbp (*i.e.* structural variation) are important sources of genetic variability in humans [23, 24]. This fact does not discount the importance of single nucleotide polymorphisms (SNPs), which were originally hailed as the dominant source of genomic variability [23]. It now appears however, that the importance of SNPs may have been initially exaggerated due to the fact that they were transparent to NGS — and thus easier to measure.

1.2 Genomic Mapping

Two alternate strategies have been proposed to overcome the weaknesses of NGS in pursuit of cheap *de novo* genome assemblies. The first focuses directly on "silver bullet" sequencing techniques that eliminate the need for large ensembles and provide much longer read lengths. These technologies are based on so-called single molecule approaches using zero-mode waveguides [25] or nanopores [26]. Such methods show great promise and companies such as Pacific Biosciences have recently reported read lengths greater than 3 kbp [27]. Even with the recent success of single-molecule methods, their approach suffers from a "data deluge" problem, where vast amounts of information obtained from a complete *de novo* sequence can swamp large-scale patterns.

A second strategy attempts to complement NGS, rather than replace it, by obtaining kilobase pair scale information from single-molecules, while neglecting the specifics of the detailed sequence [28–33]. In this approach, one seeks to obtain the genomic distance between known restriction sites, which can then be used as anchor points in a sequence assembly. There are many methods to obtain these distances, which have been recently reviewed by Dorfman et al. [31]. Among these is a promising class of techniques called genomic mapping.

Just as there are many different techniques for sequencing, there are many flavors of genomic mapping. Figure 1.2 shows a schematic of one such technique called DNA barcoding, which is representative of this class of methods. The first step of the process is depicted in the upper left quadrant, where one isolates a fragment of DNA much larger than 1 kbp and labels sequence-specific restriction sites with a fluorescent marker. The molecule of DNA is then stretched so the markers can be imaged and resolved. In Figure 1.2 this is represented in the upper right quadrant, where the DNA fragment is elongated by a very small channel on the order of 50 nm wide. The stretched molecule is then imaged using fluorescence microscopy and the signal of the fluorescence intensity is processed to obtain a measure of the genomic distance between markers, as shown in the bottom right quadrant. Finally, in the bottom left quadrant, many genomic distances from an ensemble of DNA fragments are assembled into a genome map.

The genome map resulting from the DNA barcoding process in Figure 1.2 specifies the absolute locations of the binding sites of a given restriction enzyme for an entire

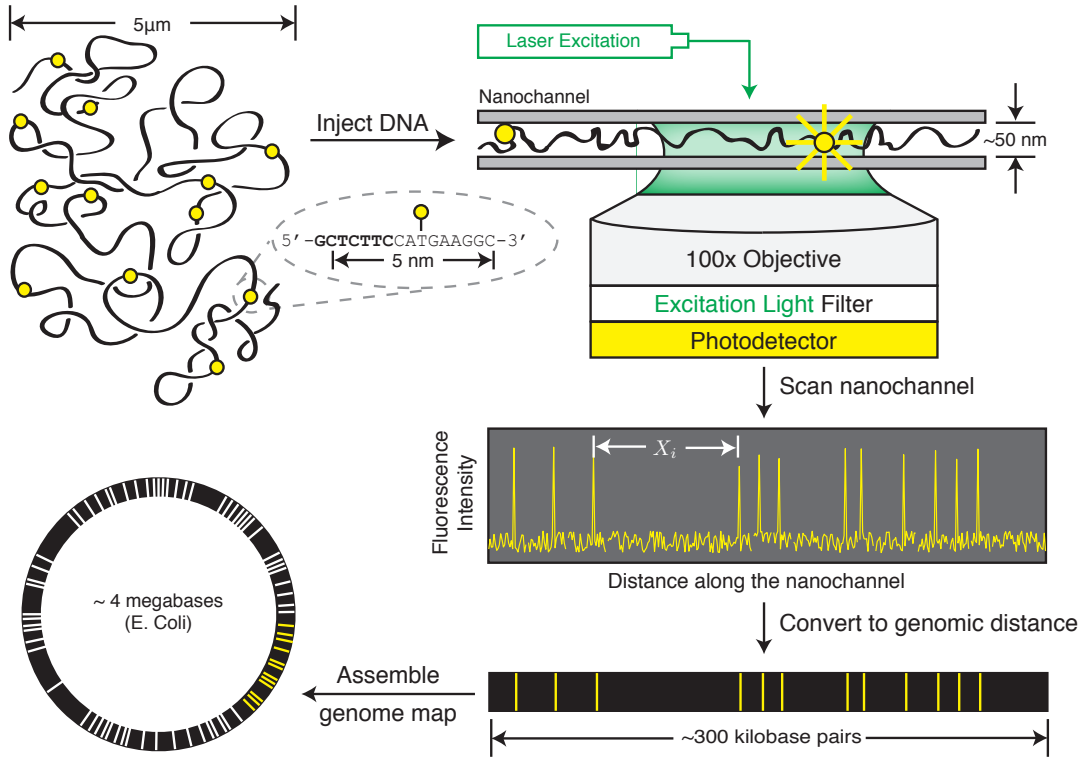

Figure 1.2: Schematic of the process of DNA barcoding. Image reproduced from Ref. 34.

genome. As hoped, these maps can be extraordinarily useful as scaffolds for sequence assembly and in directly assessing structural variation of a whole genome. Indeed, genome maps provide an invaluable tool to solve the sequencing read-length problem, which is crucial for an increased understanding of the genotype–phenotype relationship. As such, a number of interesting and useful studies involving the genomes of pathogens, agricultural products and humans have been completed using genomic mapping and are highlighted in Table 1.1. As a further testament to the usefulness of genomic mapping, several processes have been recently commercialized by the companies OpGen, Genomic Vision, PathoGenetix and BioNano Genomics.

Despite a growing number of applications, the most important genomes — humans and plants — are the most complex and remain challenging for genomic mapping technologies. Table 1.1 gives circumstantial evidence for this, showing an evolution from

smaller bacterial genomes in the early 2000's to more complicated genomes such as the characterization of wheat by BioNano Genomics in 2013 [10]. To be able to efficiently map human and plant genomes, mapping techniques need to measure genomic distances between restriction sites with (i) greater accuracy using (ii) smaller molecular ensembles consisting of (iii) larger DNA fragments with (iv) higher throughput. The method used to stretch the molecule before imaging is the key to all of these design goals.

To demonstrate the crucial role molecular stretching plays in genomic mapping, it is instructive to briefly review the ways in which the mapping technologies encompassed in Table 1.1 accomplish this task. The first-generation methods originated with "optical mapping", pioneered by the Schwartz group at Wisconsin [48]. In this class of methods, DNA is elongated on a surface using a fluid flow technique labeled molecular combing, and then restriction enzymes are added to cleave the molecules. The resulting restriction fragments are then imaged and sized.

The Wisconsin group made major advancements to this technique throughout the late 1990's and 2000's, including improvements to molecular alignment [49], labeling [50, 51] and algorithms [52]. These improvements culminated with the integration of molecular combing techniques onto a microfluidic device [53] to create a highly automated process capable of producing maps of genomes of greater than 10^6 bp. Subsequently, this technology became the basis of the genome mapping company OpGen. Other molecular combing techniques have been commercialized as well, including the company Genomic Vision, co-founded by Aaron Bensimon.

Rather than stretching DNA on surfaces, in the mid-2000's the Austin group at Princeton began experimenting with confinement as a method to elongate DNA [54, 55]. Confinement offers a number of advantages over surface fixing techniques, since the DNA analyte can remain in solution and measurements can be taken at equilibrium. In principle, greater accuracy can be obtained (or smaller ensembles can be used) by taking multiple measures of individual molecules as they fluctuate about their natural equilibrium state. (Although it is worth noting that the current commerical technology does not use this method). In addition, molecular alignment is facile since it is brought about naturally by the confining walls. Throughput can also be improved, since electrophoresis or fluid flow can bring multiple batches of DNA through the device in rapid succession. With these advantages, confinement stretching seemed like an ideal platform

Table 1.1: Notable uses of genomic mapping.

Sequence Finishing			
Year	Organism	Area of Interest	Ref.
1999	Plasmodium falciparum	malaria	35
2001	E. coli O157:H7	pathogen	36
2002	Yersinia pestis	bubonic plague	16
2003	Rhodobacter sphaeroides	photoheterotropic	37
2004	Leishmania major	pathogenic	38
Genomic Variability			
Year	Organism	Area of Interest	Ref.
2000	Homo sapien, Chr. 21	copy number variation	39
2000	Homo sapien, Y Chr.	dense repeats	40
2004	Shigella flexneri, Yersinia pestis, E. coli	structural variation	41
2005	Rhodospirillum rubrum	biofuel production	42
2007	Xenorhabdus nematophila, Xenorhabdus bovienii	model pathogen	43
2007	Oryza sativa	rice	8
2008	E. coli	methylation profiles	44
2009	Zea mays ssp. mays L	maize	9
2009	Mycobacterium avium ssp. paratuberculosis	Johne's disease	45
2010	Homo sapien	structural variation	46
2012	Homo sapien	structural variation	47
2013	Aegilops tauschii	wheat	10

for the production of complicated human or plant genomic maps.

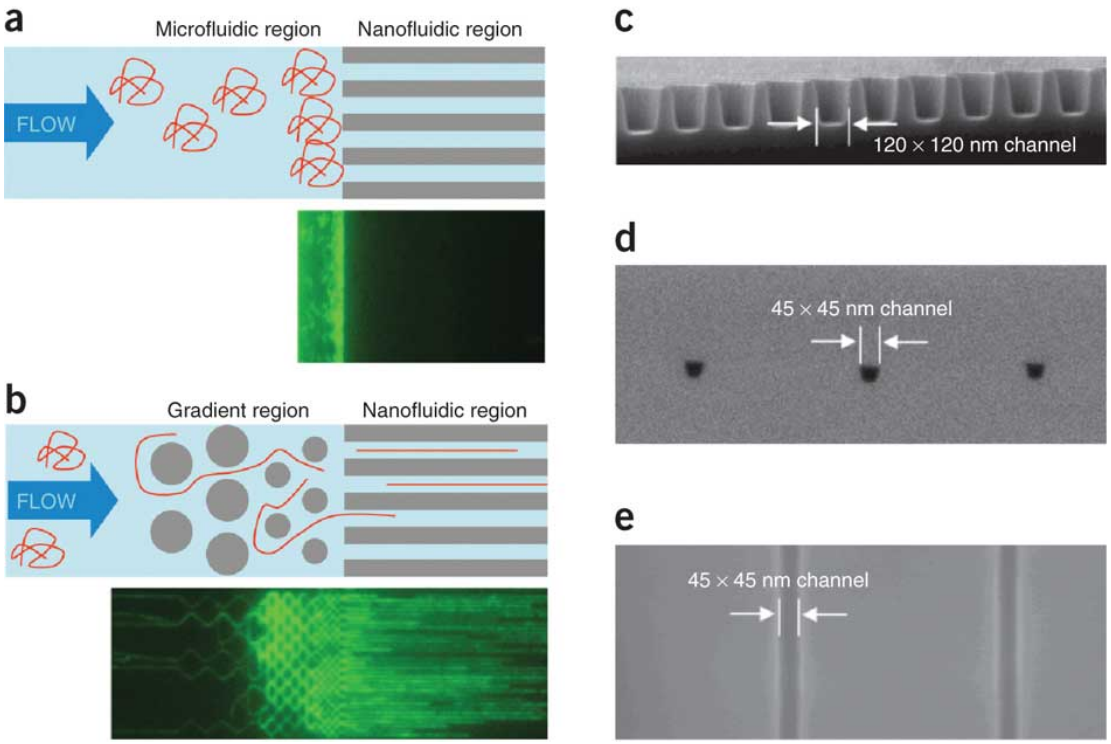

Figure 1.3: DNA entry into a nanochannel. Panel (a) shows clogging when DNA is introduced to a microchannel-nanochannel interface. Panel (b) shows DNA entering nanochannels when posts are used to pre-stretch the molecules prior to the encounter with the interface. Panels (c-e) show electron micrographs of the channels as fabricated. Figure reproduced from Ref. 47.

Nevertheless, confinement techniques had to overcome a number of technical challenges to successfully compete with molecular combing. To highlight these challenges, Figure 1.3 shows a series of fabricated nanochannels, similar to what is used in DNA barcoding outlined in Figure 1.2. The principal difficulty for confinement stretching methods concerns the extraordinary small size of the dimensions needed to stretch the DNA molecule, as shown in panels (c-e) of Figure 1.3. Because such small channels have to be used, DNA loading is a problem, and the naive attempt shown in panel (a) leads to a large amount of clogging at the inlet of the channel. This was overcome by Cao et al. [54], who used a series of posts shown in panel (b) to pre-stretch the DNA before

its introduction into the channel. In addition, the entropy gradient was further reduced at the channel entrance by slowly raising the floor, making the post-region more slit-like. Further improvements in labeling technology, enabling barcoding [51, 56] instead of in situ restriction chemistry [57], also vastly improved the accuracy and throughput of genomic mapping by confinement. Having overcome these challenges, and because of its inherent advantages, confinement-stretching genomic mapping has also been commercialized and forms the core technology for the company BioNano Genomics.

1.3 Research Outline

Notwithstanding early successes, confinement induced stretching for genomic mapping remains a new technology that is not completely understood. Indeed, DNA confinement in nanochannels and nanoslits results in a remarkable diversity of physical behavior due to the underlying relationships between polymer physics, hydrodynamics, electrostatics, confinement geometry, and surface interactions. Experimental techniques such as fluorescence microscopy, UV, e-beam and nano-imprint lithography, and optical tweezing have indeed enabled researchers to access the necessary length scales to manipulate and analyze single molecules of DNA [30, 31]. However, due to the complexity of the parameter space, the cost of generating devices, and lack of other experimental capabilities, it has been difficult for experimentalists to fully explore device properties by trial-and-error. Theorists have faced equally difficult challenges, and an analytical treatment is only possible in the simplest of cases.

As such, the goal of this dissertation is to develop a numerically tractable, statistical mechanical model capable of quantitatively reproducing the behavior of DNA in confinement. Ultimately, it is hoped that such a model can be used as a tool for design and discovery in the engineering of genomic mapping devices. However, because the physics of confined, stiff biopolymers like DNA remains poorly understood, validating such a model is not a trivial matter. Therefore, a concurrent goal of the dissertation is to explore the fundamental physics of confined DNA with coarse-grained models. Fortunately, both goals are complementary, since the exploration and discovery of interesting physical behavior in confined DNA may also play a practical role in designing new processes for genomic mapping that take advantage of such phenomena.

In line with these goals, **Chapter 2** reviews the current state of the literature surrounding the physics of DNA in confinement. This chapter introduces a coarse-grained model for DNA — the wormlike chain — and discusses analytical results for the equilibrium and dynamic properties of unconfined, dilute solutions. The polyelectrolyte nature of DNA is also discussed in this context. Chapter 2 then proceeds with a presentation of classic theoretical results by Odijk [58] and Daoud and de Gennes [59] regarding the confinement regimes of wormlike polymers. These results are then shown in the context of experimental results by the Princeton group [60] and others, and outstanding research questions are highlighted.

Having outlined the relevant theoretical and experimental results, **Chapter 3** reviews our numerical model and simulation methodology. Specifically, this chapter introduces Monte Carlo simulations and discusses both the Metropolis and chain growth methodologies used in later chapters. Chapter 3 also defines a numerical Kirkwood approximation, which allows us to estimate dynamic properties from equilibrium ensemble data. In Kirkwood's formalism continuum hydrodynamics are required for an accurate estimation of polymer dynamic properties. As such, we outline a computational fluid dynamics method that can be used to obtain the hydrodynamics of DNA in confinement.

Using these methods, **Chapter 4** addresses the properties of free solution DNA and focuses on the role of DNA as a model polymer. Here, the details surrounding the cross-over from wormlike-specific to universal polymeric behavior has been the subject of debate and confusion even for the simple case of a dilute, unconfined wormlike chain. We compute the polymer size, form factor, free energy and Kirkwood diffusivity for unconfined wormlike chains as a function of molecular weight, focusing on persistence lengths and effective widths that represent single-stranded and double-stranded DNA in a high ionic strength buffer. From our calculations, we find that very large DNA chains ($\approx$ 1,000,000 base pairs depending on the choice of size metric) are required to reach flexible, swollen non-draining coils. Furthermore, our results indicate that the commonly used model polymer λ-DNA (48,500 base pairs) does not exhibit "ideal" scaling, but exists in the middle of the transition to long-chain behavior. We subsequently conclude that typical DNA fragments used in experiments are too short to serve as an accurate model of long-chain, universal polymer behavior.

Chapter 5 shifts attention to confined chains and re-examines the nanochannel

scaling regimes outlined in chapter 2. Using large scale PERM simulations of asymptotically long wormlike chains, we show that there is in analogy to the rod-to-coil transition for unconfined wormlike polymers. Accordingly, for channel confined polymers there exists a universal, Gauss-de Gennes regime that connects the classic Odijk and de Gennes regimes of channel-confined chains. For DNA, this Gauss-de Gennes regime spans practically the entire experimentally relevant range of channel sizes, including the nanochannels used in genome mapping technology.

In light of this new regime, **Chapter 6** discusses the diffusion of DNA confined in nanochannels. Indeed, the classic results of de Gennes and Odijk describe the mobility of a semiflexible chain confined in a nanochannel only in the limits of very weak and very strong confinement, respectively. Using Monte Carlo sampling of the Kirkwood diffusivity with full hydrodynamic interactions, we show that the mobility of a semiflexible chain exhibits a broad plateau as a function of extension before transitioning to an Odijk regime. Furthermore, we observe that the width of the plateau depends on the anisotropy of the monomers. For the particular case of DNA in a high ionic strength buffer, which has highly anisotropic monomers, we predict that this Rouse-like behavior will be observed over most of the measurable chain extensions seen in experiments.

Having examined the diffusivity, **Chapter 7** further elaborates on dynamic properties and explores the relaxation time of DNA in nanochannels using a mapping between a Rouse dumbbell model and fine-grained Monte Carlo simulations. The relaxation time thus obtained agrees quantitatively with experimental data [60] using only a single $O(1)$ fitting parameter to account for the uncertainty in model parameters. In addition to validating our mapping, this agreement supports our previous estimates of the friction coefficient of DNA confined in a nanochannel in Chapter 6, which have been difficult to validate due to the lack of direct experimental data. Furthermore, the model calculation shows that as the channel size passes below approximately 100 nm (or roughly the Kuhn length of DNA) there is a dramatic drop in the relaxation time. Inasmuch as the chain friction rises with decreasing channel size, the reduction in the relaxation time can be solely attributed to the sharp decline in the fluctuations of the chain extension. Practically, the low variance in the observed DNA extension in such small channels has important implications for genome mapping.

Chapter 8 looks beyond nanochannels and examines a comprehensive theory for

a wormlike polymer strongly confined in a nanoslit. By drawing inspiration from the existing literature on ideal wormlike chains in slits and Daoud and de Gennes' idea of mapping a slit-confined chain to a two-dimensional chain, we postulate that the chain can be quantitatively described as a two-dimensional wormlike chain with a weak perturbation in the confining dimension due to deflection segments. By incorporating the effects of real chains, where the variable slit depth adds additional subtlety due to concomitant changes in the strength of excluded volume interactions, our theory predicts the existence of three distinct subregimes. We investigate the validity of our claims by performing Monte Carlo simulations of a slit-confined wormlike chain using an off-lattice implementation of the pruned-enriched Rosenbluth method. From these simulations, we find strong numerical evidence supporting our predictions, including the existence of subregimes within the Odijk regime.

Finally, **Chapter 9** offers concluding thoughts and possible directions for future research. In particular, this chapter highlights the possibilities for practical use of our coarse-grained model as an engineering tool for genomic mapping devices.

Chapter 2

Stretching DNA in Confinement

2.1 Introduction

As stated in Chapter 1, our ultimate goal is to develop a statistical mechanical model capable of describing the physical properties of a large fragment of double-stranded DNA (dsDNA) in a nanoscale genomic mapping device. Unfortunately, from a computational perspective, an all-atom representation of dsDNA is intractable for genomic lengths greater than a few hundred base pairs. This is of course far too short for our purposes. In addition, such a model suffers from a large number of detailed parameters, which cloud the physical intuition we hope to attain. In this matter we agree with J. W. Gibbs who said, "One of the principal objects of theoretical research ... is to find the point of view from which the subject appears in its greatest simplicity [61]."

We therefore turn to a coarse-grained representation of dsDNA, where the properties of the atoms are aggregated in some way to obtain a model with larger inherent length and time scales. From this perspective, the so-called wormlike chain (WLC) or Kratky-Porod model [62] is a good choice for a coarse-grained model of DNA in nanoscale confinement. The WLC model is a homogeneous polymer model that models DNA as a sequence independent chain with a minimum resolution of about 10 base pairs. This greatly reduces the degrees of freedom from an all-atom representation, and assumes that specific base-pair interactions can be averaged out at the intrinsic length scales in the model. Fortunately, at the contour lengths of interest in genomic mapping ($\geq$ 1 kbp), this loss of detail is unimportant, and the sequence dependence of DNA plays a

negligible role. Furthermore, since the minimum length scale in the WLC model is only a few nanometers, it is fine-grained enough to capture the nanoscale effects necessary for modeling genomic mapping.

In addition to occupying the sweet spot of length and time scales, the wormlike chain model is an excellent description of DNA physics. This is primarily due to the inherent stiffness of the DNA backbone and was experimentally validated by a series of direct mechanical elasticity measurements in the 1990's [63, 64]. Nevertheless, as a neutral polymer model, the WLC model cannot directly account for the polyelectrolyte nature of DNA. Indeed, DNA is a highly charged species and experiences a number of phenomena, such as electrophoresis, that cannot be accounted for by the WLC model alone. While this is an inherent limitation, a neutral model is sufficient for our purposes in assessing the change in polymer size upon confinement. It does complicate the picture however, and renders the WLC model parameters a function of ionic strength [65, 66].

For the remainder of this chapter, we will review some important theoretical results for wormlike chains and the accompanying experimental data for DNA in both free solution and confinement. We begin by defining the wormlike chain model and highlight some of its important properties in free solution. In particular, we are interested in obtaining a prediction for the size of the chain and the dependence of this size on the model parameters. We are also interested in the dynamics of the polymer — specifically the diffusivity. We then describe wormlike chains in confinement, first in channels and then in slits. Here, we are again interested in the size of the molecule and its interaction with the confining geometry. In particular, we would like to understand the physics that drives chain stretching, since genomic mapping applications seek to exploit this property in particular.

2.2 Wormlike Chains at Equilibrium

The wormlike chain model is schematized in Figure 2.1. Here, the polymer is defined by a set 3-dimensional vectors $\{\mathbf{r}_i\}$, $i \in [1, N_b]$ of "bead" positions, connected by rigid-rods of length a. The resulting polymer has a contour length $L = (N_b - 1)a$. Note that the contour length of 1 kbp of DNA is about 0.3μm, using a standard heuristic that 1bp contributes about 0.34 nm to the chain length [67].

Like all statistical mechanicals model, the wormlike chain is defined by a Hamiltonian, $H(\{\mathbf{r}_i\})$, which can be used to compute the phase space probability density function

$$\psi(\{\mathbf{r}_i\}) = \frac{\exp[-\beta H(\{\mathbf{r}_i\})]}{Z} \tag{2.1}$$

where β is the inverse Boltzmann factor $(k_B T)^{-1}$ and Z is the configurational partition function

$$Z = \int \exp[-\beta H(\{\mathbf{r}_i\})] d\{\mathbf{r}_i\} \tag{2.2}$$

Using ψ and Z and the tools of modern equilibrium statistical mechanics, we can compute all of the properties of bulk thermodynamic properties of interest.

The WLC Hamiltonian includes two terms

$$H(\{\mathbf{r}_i\}) = U_{\text{bend}}(\{\mathbf{r}_i\}) + U_{\text{ev}}(\{\mathbf{r}_i\}) \tag{2.3}$$

where the first term on the right hand side accounts for a bending potential and the second term gives an excluded volume interaction between beads of strength w. For the sake of simplicity — especially when considering polymer dynamics — we only discuss a discrete wormlike chain model in this chapter. However, it is also possible to describe a continuous wormlike chain models, and several are discussed in the literature [68–70].

The local stiffness of the WLC chain distinguishes it from other homogeneous polymer models. This stiffness arises from the bending potential [71]

$$U_{\text{bend}} = \kappa \sum_{i=1}^{N_b-2} (1 - \mathbf{u}_i \cdot \mathbf{u}_{i+1}) \tag{2.4}$$

where κ is the bending constant and $\mathbf{u}_i = a^{-1}(\mathbf{r}_{i+1} - \mathbf{r}_i)$ is a unit bond vector. This bending potential causes the path of the polymer to have a persistent memory, which can be characterized by an autocorrelation function of the tangent vectors along the backbone of the polymer

$$C(s) = \langle \mathbf{u}(s) \cdot \mathbf{u}(0) \rangle \tag{2.5}$$

where $s = a(i-1)$ for $i \in [1, N_b]$. Using the bending potential in Eq. 2.4 to evaluate the autocorrelation function in Eq. 2.5, we obtain an exponential decay [68, 72]

$$C(s) = \exp(-s/l_p) \tag{2.6}$$

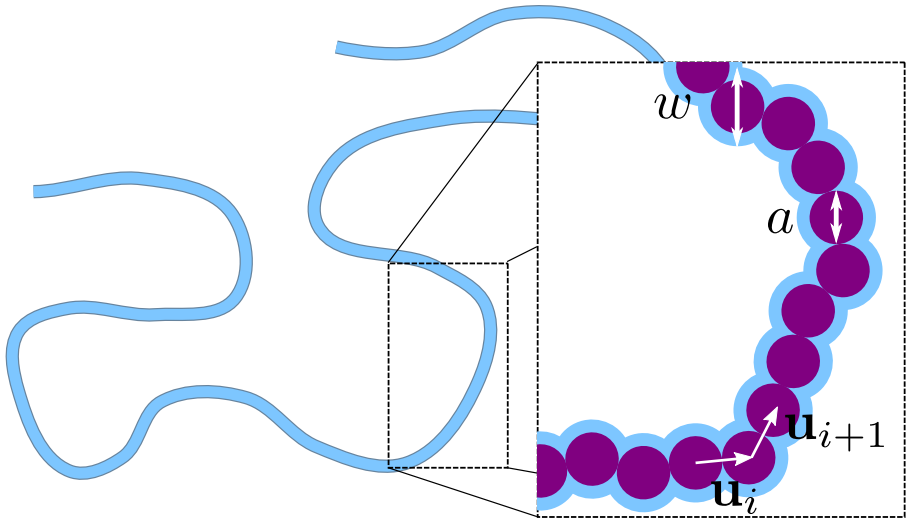

Figure 2.1: Schematic of a wormlike chain in free solution. This discrete version of the model is composed of N_b osculating beads of size a. In addition, there is a bending potential between nearest neighbor bonds and a hard-core excluded volume interaction.

with a characteristic length, l_p. This length is called the persistence length, and it is related to the bending constant by

$$l_p \approx \frac{\kappa}{k_B T} \tag{2.7}$$

The persistence length for DNA is about 50 nm, but this value depends somewhat on solvent conditions and is discussed in greater detail in Chapter 4.

The size of the DNA molecule in free solution can be described by the root-mean-squared (rms) end-to-end distance, R. Since the molecule's conformational statistics are rotationally invariant, the averaged end-to-end distance is zero, but the rms value is well defined. For a wormlike chain, the square end-to-end distance is given by [68, 72]

$$R^2 = 2Ll_p \left[1 - \frac{l_p}{L}\left\{1 - \exp\left(-\frac{L}{l_p}\right)\right\}\right] \tag{2.8}$$

When the chain is short compared to the persistence length, $L \ll l_p$, the DNA is essentially a rod-like molecule whose size is $R \approx L$. By contrast, long chains where $L \gg l_p$ lead to the end-to-end distance

$$R^2 \approx 2Ll_p \tag{2.9}$$

The radius of gyration is another measure of polymer size and is directly measureable in x-ray and neutron scattering experiments. The radius of gyration for an ideal wormlike chain is [72, 73]

$$R_g^2 = Ll_p \left[\frac{1}{3} - \frac{l_p}{L} + 2\frac{{l_p}^2}{L^2} - 2\frac{{l_p}^3}{L^3} \left\{ 1 - \exp\left(-\frac{L}{l_p} \right) \right\} \right] \tag{2.10}$$

In the polymer physics literature [72, 74], one often refers to a statistical segment length, b. The latter length scale originates from the concept of an equivalent random walk, which has an end-to-end distance

$$R = b\sqrt{N} \tag{2.11}$$

By comparing 2.11 and 2.9, and by using the definition $N \equiv L/b$, we obtain the relationship between the statistical segment length (also known as the Kuhn length) and the persistence length, $b = 2l_p$.

Until now, we have neglected the role of excluded volume (EV) interactions on the behavior of wormlike chains. Excluded volume interactions are due to intrachain interactions between distal segments of the chain and are the second term in the Hamiltonian in Eq. 2.3. The excluded volume potential is described by [69, 75]

$$U_{\text{ev}} = \sum_{i=1}^{N_b} \sum_{j=1}^{i-1} \begin{cases} \infty, & r_{ij} \leq w \\ 0, & r_{ij} > w \end{cases} \tag{2.12}$$

where $r_{ij} = |\mathbf{r}_j - \mathbf{r}_i|$ is the distance between two polymer beads. The strength of the excluded volume interactions is characterized by the width, w, of the hard-core repulsion in Eq. 2.12. This gives rise to a second virial coefficient for the Kuhn monomer of wb^2. For DNA the effective width of the molecule, comprising both steric and electrostatic forces, is about 5 nm.

The polymer size is altered because of excluded volume interactions, and Eq. 2.8 is inadequate for real chains. Note that we use the terminology "real chain" and "ideal chain" analogously to "real gas" and "ideal gas" to indicate chains with and without intrachain interactions respectively. Flory developed an intuitive theory for the size of a chain with excluded volume, which has been uncreatively labeled "Flory theory" [69, 72, 76]. In Flory theory the free energy is expressed as a sum of two terms

$$\frac{F}{k_B T} \sim \frac{R^2}{Ll_p} + N^2 \frac{l_p^2 w}{R^3} \tag{2.13}$$

where the first corresponds to the entropic penalty to stretch the chain and the second gives a penalty for chain overlap. Finding the minimum free energy with respect to R gives

$$R \sim l_p \left(\frac{w}{l_p}\right)^{2\nu-1} \left(\frac{L}{l_p}\right)^{\nu} \tag{2.14}$$

with $\nu = 3/5$ being the Flory exponent. Flory theory contains a number of subtle errors [69, 77–79], but predicts a value remarkably close to the Flory exponent obtained by more sophisticated modern methods, $\nu = 0.5876$ [80, 81].

Because excluded volume effects are due to interactions between chain segments, the sum total of their magnitude depends on the length of the chain. This is because it is increasingly likely that distal chain segments will interact as the chain gets longer, causing the molecule to swell relative to the case where EV is absent. Consequently, very short chains exhibit negligible excluded volume interactions, while long chains experience important effects. The contour length at which excluded volume becomes order k_BT is called the thermal blob length, l_{T}. A rearrangement of Eq. 2.14

$$R \sim (Ll_p)^{1/2} \left[\frac{w}{l_p}\left(\frac{L}{l_p}\right)^{1/2}\right]^{2\nu-1} \tag{2.15}$$

reveals the thermal blob length

$$\frac{l_{\mathrm{T}}}{l_p} \sim \frac{w}{l_p}\left(\frac{L}{l_p}\right)^{1/2} \tag{2.16}$$

as the length scale which divides the crossover from ideal chain scaling in Eq. 2.9 to real chain scaling in Eq. 2.14. The dimensionless quantity on the right hand size of Eq. 2.16 is also sometimes called the excluded volume parameter, z.

While R is a useful theoretical construct, the radius of gyration, R_g, is more prevalent since it can be measured from a number of experimental methods [74]. The radius of gyration measures the rms distance between parts of the molecule and its center of mass. Figure 2.2 presents a compilation of experimental data for the radius of gyration as a function of the contour length of dsDNA for a wide range of experimental conditions [82–108]. As seen in the figure, the radius of gyration predicted by the wormlike chain model is in excellent agreement with the experimental measurements. We can also clearly see the different regimes of DNA size in this figure. For short chains, the radius of gyration

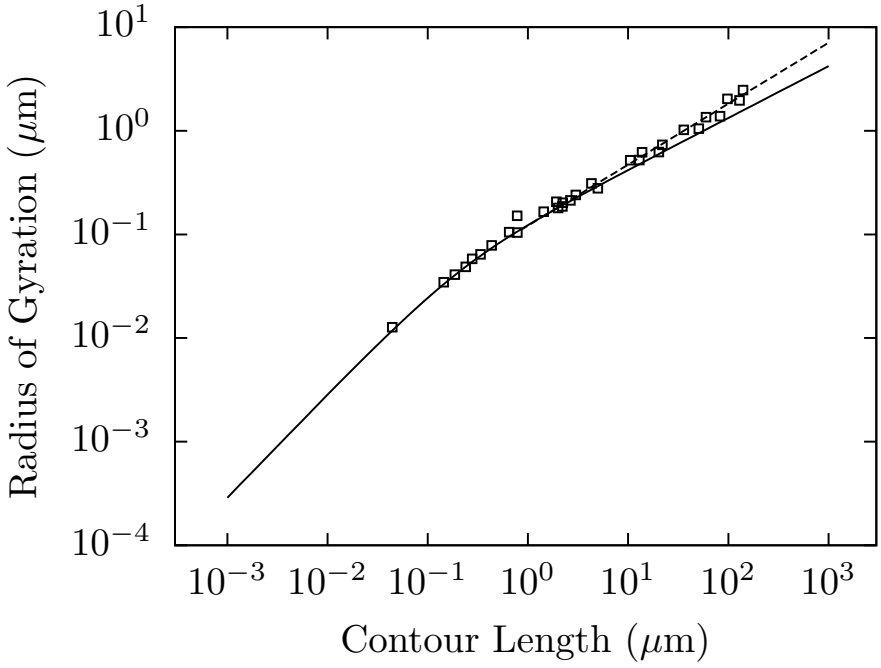

Figure 2.2: Radius of gyration (squares) of DNA for a wide range of experimental conditions found in the literature [82–108]. The solid line is Eq. 2.10 and the dashed line shows the N^{ν} scaling.

increases linearly with the contour length. For moderate values of the contour length, the chain is an ideal random walk with the scaling in Eq. 2.11. At the largest contour lengths, the radius of gyration breaks away from the ideal chain scaling and begins to follow the self-avoiding random walk scaling in Eq. 2.14.

Finally, the introduction of the effective width w into the wormlike chain model brings to light an important ratio

$$\alpha = \frac{b}{w} \tag{2.17}$$

called the (Kuhn) monomer anisotropy [109]. The monomer anisotropy represents a competition between the effects of stiffness and excluded volume on the polymer. For DNA, α is about 20 in commonly used buffers, indicating that stiffness dominates excluded volume interactions on the monomer length scale. This fact plays an important role in the equilibrium behavior of DNA and is one of the properties that distinguishes DNA from typical synthetic polymers.

2.3 Wormlike Chain Dynamics

Until now, we have focused solely on equilibrium properties of wormlike chains. For dynamic quantities, Kirkwood and Riseman developed a formalism treating the solvent as an implicit continuum, leaving only the polymer degrees of freedom [110–113]. This can be expressed quite generally as a Fokker–Planck equation for the time evolution of the probability density of the chain configuration [114]

$$\frac{\partial \psi}{\partial t} = \sum_{i=1}^{N_b} \sum_{j=1}^{N_b} \frac{\partial}{\partial \mathbf{r}_i} \cdot \boldsymbol{\mathcal{D}}_{ij} \cdot \left[\frac{\partial \psi}{\partial \mathbf{r}_j} + \frac{\psi}{k_B T} \frac{\partial H}{\partial \mathbf{r}_j} \right] \tag{2.18}$$

where $\boldsymbol{\mathcal{D}}_{ij}$ is a 3×3 diffusion tensor. It is hoped that one can obtain dynamic properties of interest to genomic mapping from Eq. 2.18, which includes the diffusion coefficient of the center of mass of the polymer.

One of the distinguishing features of Eq. 2.18 is its capacity to incorporate hydrodynamic interactions (HI), which are a crucial part of correctly modeling polymer dynamics. HI are depicted in Figure 2.3. Here, one part of the chain is perturbed by a force F and induces a flow in the solvent. This flow is long-ranged and entrains other parts of the chain by an induced drag force. Thus, HI are solvent mediated intrachain interactions between distal segments of the chain and are the dynamic analogue of excluded volume. However, it is important to recognize that HI are a dynamic effect only. Because the time-averaged force on the polymer from the solvent is zero, HI do not alter ψ in the long time limit.

Hydrodynamic interactions enter Eq. 2.18 through the diffusion tensor. To find this tensor, consider the change in velocity of the solvent, $\mathbf{v}_i$, from a quiescent state, due to the drag $\mathbf{F}_i$ on the chain

$$\mathbf{v}_i = \sum_{j=1}^{N_b} \boldsymbol{\Omega}_{ij} \cdot \mathbf{F}_j \tag{2.19}$$

The hydrodynamic tensor in this expression is the 3×3 Oseen–Burgers tensor from the solution of the Stokes equation due to a point force perturbation [114],

$$\boldsymbol{\Omega}_{ij} = \frac{1}{8\pi\eta r_{ij}} \left(\mathbf{I} + \frac{\mathbf{r}_{ij}\mathbf{r}_{ij}}{r_{ij}^2} \right) \text{ for } i \neq j \tag{2.20}$$

In Eq. 2.20, $\mathbf{r}_{ij}$ is the vector $\mathbf{r}_j - \mathbf{r}_i$, the magnitude is r_{ij} and η is the solvent viscosity. The drag force is given by the product of the bead friction coefficient, ζ, and the difference

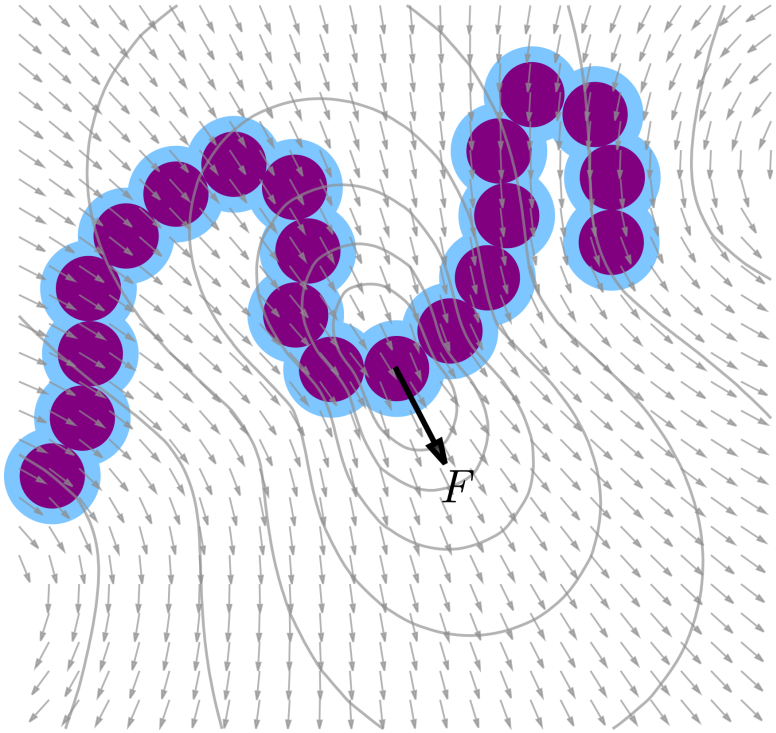

Figure 2.3: Schematic of hydrodynamic interactions for a discrete polymer chain. A point force is applied to one of the beads and the resulting flow field affects the dynamics of the rest of the chain.

between the bead velocity $\mathbf{u}_i$ and the solvent velocity,

$$\mathbf{F}_i = \zeta \left(\mathbf{u}_i - \mathbf{v}_i \right) \tag{2.21}$$

Substituting Eq. 2.19 into Eq. 2.21 gives

$$\mathbf{F}_i = \zeta \mathbf{u}_i - \zeta \sum_{j=1}^{N_b} \mathbf{\Omega}_{ij} \cdot \mathbf{F}_j \tag{2.22}$$

which is a self-consistent equation for the drag force in terms of the bead velocities and the hydrodynamic tensor. Defining the diffusion tensor according to the fluctuation-dissipation theorem

$$\mathbf{u}_i = \frac{1}{k_B T} \sum_{j=1}^{N_b} \boldsymbol{\mathcal{D}}_{ij} \cdot \mathbf{F}_j \tag{2.23}$$

and rearranging Eq. 2.22 yields the sought-after expression

$$\boldsymbol{\mathcal{D}}_{ij} = k_B T \left(\frac{\delta_{ij}}{\zeta} \mathbf{I} + \boldsymbol{\Omega}_{ij} \right) \tag{2.24}$$

It is worth noting that in the general formulation of the Kirkwood–Riseman theory, Eq. 2.24 is not exact when constraints on the system (*e.g.* rigid bond angles) are included [115]. In this case, Eq. 2.24 only represents the first two terms in an expansion in the strength of the hydrodynamic interactions [116].

Eq. 2.24 completes the Kirkwood–Riseman theory in Eq. 2.18, and we can in principle calculate dynamic quantities such as the center of mass diffusion coefficient. However, solving Eq. 2.18 directly for the center of mass diffusion coefficient is an intractable problem for a chain with so many degrees of freedom [117, 118]. Accordingly, a number of approximations are commonly made to arrive at a concrete prediction for the diffusivity. The most mild of these involves a straightforward numerical solution. The numerical calculations are performed by reformulating Eq. 2.18 in terms of a set of coupled stochastic differential equations and integrating in time akin to Molecular Dynamics [117, 119]. Such calculations are commonly referred to as Brownian Dynamics simulations (with hydrodynamics) and can be quite computationally expensive.

In lieu of an expensive Brownian Dynamics calculation, we follow Kirkwood and Riseman in making a so-called pre-averaging approximation for the center of mass diffusivity [110, 116]. The center of mass diffusion tensor is defined as

$$\mathbf{u}_{\mathrm{cm}} = \frac{1}{k_B T} \boldsymbol{\mathcal{D}}_{\mathrm{cm}} \cdot \mathbf{F}_{\mathrm{cm}} \tag{2.25}$$

where

$$\mathbf{u}_{\mathrm{cm}} = \frac{1}{N_b} \sum_{i=1}^{N_b} \langle \mathbf{u}_i \rangle \tag{2.26}$$

and

$$\mathbf{F}_{\mathrm{cm}} = \sum_{i=1}^{N_b} \langle \mathbf{F}_i \rangle \tag{2.27}$$

Summing Eq. 2.23 over i, dividing by N_b and taking the ensemble average gives

$$\mathbf{u}_{\mathrm{cm}} = \frac{1}{N_b k_B T} \sum_{i=1}^{N_b} \sum_{j=1}^{N_b} \langle \boldsymbol{\mathcal{D}}_{ij} \cdot \mathbf{F}_j \rangle \tag{2.28}$$

which is as far as we can progress without an approximation. The pre-averaging approximation consists in separating $\mathbf{F}_j$ from $\boldsymbol{\mathcal{D}}_{ij}$ in the ensemble average

$$\mathbf{u}_{\mathrm{cm}} = \frac{1}{N_b k_B T} \sum_{i=1}^{N_b} \sum_{j=1}^{N_b} \langle \boldsymbol{\mathcal{D}}_{ij} \rangle \cdot \langle \mathbf{F}_j \rangle \tag{2.29}$$

resulting in the well-known Kirkwood double-sum formula [111]

$$\boldsymbol{\mathcal{D}}_{\mathrm{cm}} = \frac{1}{N_b^2} \sum_{i=1}^{N_b} \sum_{j=1}^{N_b} \langle \boldsymbol{\mathcal{D}}_{ij} \rangle \tag{2.30}$$

The pre-averaging approximation in Eq. 2.29 is essentially a hydrodynamic mean-field approximation, where a chain segment feels the average HI at every point in time. Accordingly, the approximation neglects the dynamic correlations that exist between the intramolecular hydrodynamic interactions at different points in time [120, 121]. Because these correlations are also small at short times, the Kirkwood diffusivity in Eq. 2.30 can be thought of as a short-time diffusion coefficient. Using linear response theory, Fixman showed that dynamic intramolecular correlations always decrease the diffusivity, making the Kirkwood estimate an upper bound to the true diffusivity [120].

An alternative technique called the rigid-rod approximation has been tried in an attempt to overcome the limitations of pre-averaging. In this approach, one numerically solves Eq. 2.28 using the expression for the drag forces in Eq. 2.22 before performing the ensemble average using Monte Carlo data [122]. This technique does indeed circumvent the pre-averaging, but instead assumes that the dynamic correlations are between "rigid" conformations of the polymer. This approach has been questioned theoretically [123], and numerical results yield errors similar or worse than the pre-averaging approximation [124].

Finally, we are prepared to obtain estimates of $\mathcal{D}_{\mathrm{cm}}$ for a wormlike chain. Because the Kirkwood expression depends on a configurational average only, we can use the configurational statistics from Section 2.2 to evaluate Eq. 2.30. In estimating the scalar diffusion coefficient, accounting for the rotational invariance of the Oseen tensor makes another simplification to Eq. 2.30 and yields [111, 114, 116]

$$\mathcal{D}_{\mathrm{cm}} = \frac{k_B T}{3\pi\eta N_b a} + \frac{k_B T}{6\pi\eta N_b^2} \sum_{i=1}^{N_b} \sum_{j=1}^{N_b} \left\langle \frac{1}{r_{ij}} \right\rangle \tag{2.31}$$

In Eq. 2.31 we have assumed that $\zeta = 3\pi\eta a$, where a is the hydrodynamic diameter.

For wormlike chains without excluded volume, we expect two different regimes of diffusive behavior depending on the length of the chain. In Section 2.2, we showed that a chain shorter than l_p behaves similar to a rigid rod. In this case, the configurational statistics give $\langle R \rangle^{-1} \sim L^{-1}$. Substituting this into Eq. 2.31 and approximating each sum by an integral over the chain length gives [116],

$$\mathcal{D}_{\text{cm}} \sim \frac{k_B T}{\eta L} \ln\left(\frac{L}{a}\right) \tag{2.32}$$

in the limit that $L \gg a$ where a is the hydrodynamic diameter of the rod.

Longer chains are more flexible and form gaussian coils. In Section 2.2, we found that the size of the coil scales like $\langle R \rangle^{-1} \sim (l_p L)^{-1/2}$. Substituting this into Eq. 2.31 and making an approximation for the integrals gives [110, 116]

$$\mathcal{D}_{\text{cm}} \sim \frac{k_B T}{\eta L}\left[1 + c\left(\frac{L}{l_p}\right)^{1/2}\right] \tag{2.33}$$

where c is a constant.

There is an interesting cross-over in Eq. 2.33 due to the strength of the hydrodynamic interactions. When the chain is short, the HI are weak, and the second term in the brackets on the right hand side of Eq. 2.33 is small. Here the diffusivity reduces to [125]

$$\mathcal{D}_{\text{cm}} \sim \frac{k_B T}{\eta L} \tag{2.34}$$

which is commonly known as Rouse diffusion. In this case, solvent velocity penetrates the entire coil, and each chain segment acts as an independent friction center. The resulting friction is proportional to the number of segments in the chain, which is reflected in the denominator of Eq. 2.34. Because of the unfettered motion of the solvent throughout the coil, a chain with a diffusion like Eq. 2.34 is sometimes called "free-draining."

By contrast, when the chain is long, HI are strong, and the second term on the right hand side of Eq. 2.33 is large. In this case the diffusivity becomes [126]

$$\mathcal{D}_{\text{cm}} \sim \frac{k_B T}{\eta (L l_p)^{1/2}} \tag{2.35}$$

which is known as Zimm diffusion. Here, hydrodynamic interactions between chain segments at the exterior of the coil screen interior segments from additional HI. Because

the interior of the coil is screened, the solvent velocity at the center is near zero and the coil is said to carry solvent with it. The fluid is not literally trapped inside the polymer coil, but the hydrodynamics make the chain appear like a solid object that cannot "drain" the fluid from its interior. As a consequence, the coil is labeled "non-draining" and the friction is proportional to the size of the coil, as seen in the denominator of Eq. 2.35.

Implementing the Kirkwood calculation in Eq. 2.31 for chains with excluded volume is considerably more difficult. Renormalization group calculations indicate that Zimm scaling holds for chains with excluded volume in the long chain limit, and that the friction remains proportional to the coil size [127]

$$\mathcal{D}_{\mathrm{cm}} \sim \frac{k_B T}{\eta l_p N^{\nu}} \tag{2.36}$$

Therefore, for most practical purposes, one may think of the Zimm result for an ideal chain and the renormalization group result for the excluded volume chain as having the same qualitative behavior, except for the scaling with respect to N. Nevertheless, there are subtleties due to the interplay of excluded volume and hydrodynamic interactions that make the exact details of this regime more subtle [128].

Figure 2.4 shows a collection of experimental data for the diffusion coefficient of dsDNA obtained in a range of experimental conditions and using a wide range of experimental techniques [82–108]. Alongside this data, we show a theory by Yamakawa [129], which was derived specifically for wormlike chains without excluded volume interactions, and we show Eq. 2.36 for chains with EV. Yamakawa's theory is too complicated to detail here, but it suffices to note that it reduces to Eq. 2.32 and Eq. 2.35 in the short and long chain limits respectively and crosses over smoothly in between. As is the case with the radius of gyration in Figure 2.2, the diffusion coefficient for a wormlike chain [129] describes the DNA experimental data very well. For longer chains the diffusivity appears to deviate from Yamakawa's theory and follows Eq. 2.36 for a non-draining coil with excluded volume.

Finally, we note that Yamakawa's theory and Eq. 2.32 require a value for the hydrodynamic diameter of DNA. For DNA, a is approximately 2 or 3 nm, which we discuss in detail in Chapter 4. In principle, the hydrodynamic diameter corresponds to the surface of shear of the molecule. However, due to the number of approximations involved in

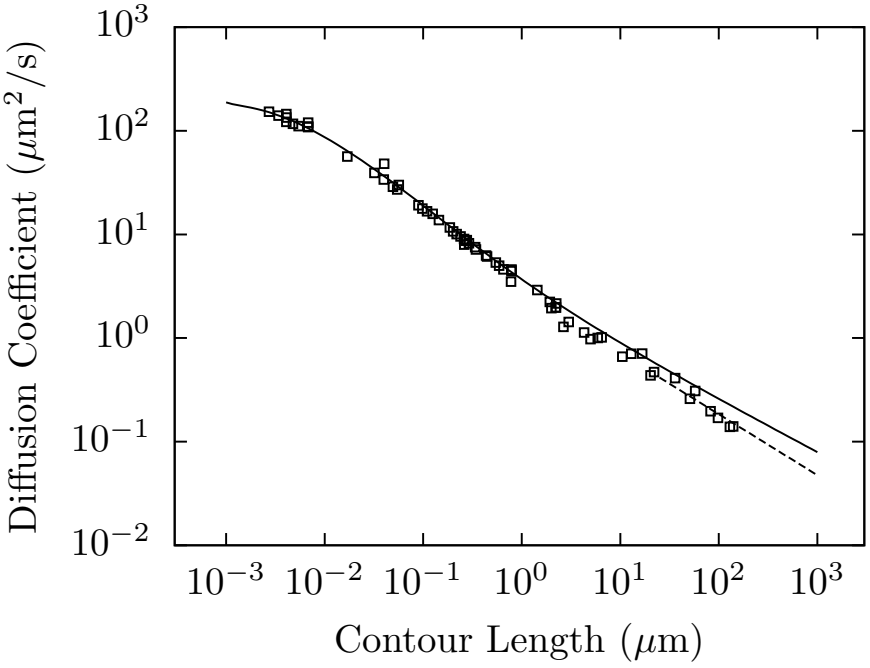

Figure 2.4: Diffusion coefficient (circles) of DNA for a wide range of experimental conditions found in literature [82–108]. The solid line represents the theory of Yamakawa [129] for the diffusion coefficient of a wormlike chain without excluded volume with a bead hydrodynamic diameter of $a = 2.28$ nm; the dashed line indicates scaling like $N^{-\nu}$.

obtaining Eq. 2.32, we prefer to think of the diameter a as a purely phenomenological parameter.

2.4 Nanochannel Confinement

Having reviewed the size and diffusion of DNA in bulk dilute solutions, we turn our attention to DNA in confinement. In this section we focus on DNA confined in a square channel of size, D, which is unbounded in the axial direction. Unlike other stretching methods, no external forces are required for DNA to extend in such a channel; the molecule is stretched in its equilibrium conformation. The principle metric we consider is the fractional extension, $\langle X/L \rangle$, which we define here as the mean span of the DNA along the channel axis relative to its contour length.

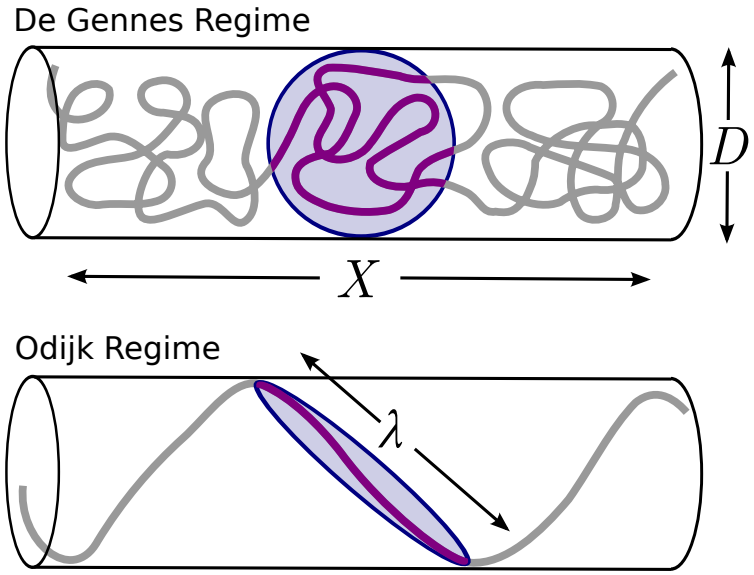

Figure 2.5: Classic regimes of confinement for DNA. When $D \ll R_g$ the molecule enters the de Gennes regime shown in the top half of the figure [59]. The de Gennes regime is characterized by the formation of blobs of size D. When $D \ll l_p$ the polymer enters the Odijk regime, shown in the bottom half of the figure [58]. The Odijk regime is characterized by the formation of deflection segments of length λ.

For most of the last thirty years, channel confinement was considered a solved problem. Classic derivations by Daoud and de Gennes [59] and by Odijk [58] categorized the effect of confinement on DNA stretching into three regimes of distinct behavior: a weak confinement regime, a de Gennes regime and an Odijk regime. The latter two regimes are illustrated in Figure 2.5.

The weak confinement regime occurs when the channel size is large compared to radius of gyration of DNA. When the chain is longer than the thermal blob length, this occurs if $D \gg (wl_p)^{1/5}L^{3/5}$, using Eq. 2.14 with $\nu = 3/5$. In this case, the polymer behaves mostly as a bulk coil, except for the loss of some translational degrees of freedom. Therefore, in the weak-confinement regime the fractional extension (like other measures of polymer size) is a function of the chain length only

$$\langle X/L \rangle \approx \frac{R_g}{L} \sim L^{-2/5} \quad \text{(Bulk)} \tag{2.37}$$

and is independent of the channel size.

As the channel size decreases further, the chain begins to be squeezed by the presence of the nearby walls and thus its fractional extension begins to increase. This regime was described in the 1970's by de Gennes and coworkers [59, 77], and is commonly referred to as the "de Gennes" regime. As shown in Figure 2.5, the chain is envisioned as a series of blobs of size D where the chain is self-avoiding within the blob.

Using the blob concept, the fractional extension of the chain can be readily obtained from a Flory theory for a 1D self-avoiding walk [71]. Here the free energy is expressed as

$$\frac{F}{k_B T} \sim \frac{X^2}{N_{\text{blob}} D^2} + N_{\text{blob}}^2 \frac{D}{X} \tag{2.38}$$

where the first term is the entropy of stretching a chain of blobs and the second term is the binary energy of interaction between two blobs. The number of blobs in the chain $N_{\text{blob}} = L/L_{\text{blob}}$ is obtained by finding the contour length L_{blob} inside a self-avoiding blob of size D

$$D = L_{\text{blob}}^{3/5} (w l_p)^{1/5} \tag{2.39}$$

Finding the minimum of Eq. 2.13 and using Eq. 2.39 to find the number of blobs yields,

$$\langle X/L \rangle \approx (w l_p)^{1/3} D^{-2/3} \quad \text{(de Gennes)} \tag{2.40}$$

The particular scaling with the channel size D comes from using the Flory exponent $\nu = 3/5$. If we use a more accurate result for the Flory exponent [80], $\nu = 0.5877$, the scaling becomes $\langle X \rangle \sim D^{-0.7015}$ [71].

Given our interest in dynamics, one can make further predictions of the diffusivity and the relaxation time in this regime [30, 55, 71, 77] Assuming that hydrodynamic interactions are strong, Brochard and de Gennes argued that the chain should be Zimm-like in confinement, leading to [71, 77, 130]

$$\mathcal{D}_{\text{cm}} = \frac{k_B T}{\eta \langle X \rangle} \tag{2.41}$$

Using this and an expression for the extension fluctuations from Eq. 2.38, Tegenfeldt *et. al.* derive

$$\tau \sim \frac{\eta}{k_B T} \frac{L^2 (l_p w)^{2/3}}{D^{1/3}} \tag{2.42}$$

for the longest relaxation time of the polymer [30, 55, 71].

In the opposite limit of very strong confinement, where $D \ll l_p$, Odijk [58] envisioned the chain as a series of deflection segments as shown in the bottom half of Figure 2.5. Deflection segments are stiff portions of the chain of length [58],

$$\lambda = D^{2/3} l_p^{1/3} \tag{2.43}$$

and describe the average distance the chain can travel before a collision with the wall. Odijk determined that the fractional extension has the non-power law form [58]

$$\langle X/L \rangle = 1 - 2\alpha (D/l_p)^{2/3} \quad \text{(Odijk)} \tag{2.44}$$

where α is a prefactor. Gommper and colleagues [131, 132] have computed the parameter α (as well as other prefactors describing the free energy and the fluctuations about the mean extension) to very high precision for both circular tubes and rectangular channels.

These theories went unchallenged until the mid-2000s, when single molecule visualizations of DNA in confinement became possible. The first experiments in nanochannels were produced by the aforementioned Princeton group in a seminal paper by Tegenfeldt *et. al.* in 2004 [55]. In this proof-of-principle exhibition, they fabricated 100 nm nanochannels using nanoimprint lithography, and successfully measured the extension of a ladder of concatamers of λ-DNA. Making 20 measurements in a 100 nm channel over the course of a minute, they measured the fractional extension ($\langle X/L \rangle = 0.38$) with a resolution of about 150 nm (400 bp) and the relaxation time (≈ 1.6s) [55].

Reisner *et. al.* went beyond beyond the proof of principle stage and fabricated an impressive range of channel sizes to probe the physics of confined DNA more deeply. Subsequently, they measured the fractional extension (up to ≈ 0.75) and the extension relaxation time as a function of D [60], which are shown in Figure 2.6(a) and (b) respectively. Surprisingly, their results disagree with the scaling laws predicted by de Gennes. For the fractional extension, they measured a slope of -0.85 instead of $-2/3$ (Eq. 2.40), and for the relaxation time they obtained a slope of -0.9 instead of $-1/3$ (Eq. 2.42). The experimental results for the extension have since been corroborated with better statistics using tapered nanochannels [133]. In tapered nanochannels the degree of confinement is slowly and continuously changed slowly along the channel axis, allowing for much better sampling of the fractional extension on a single nanofluidic device.

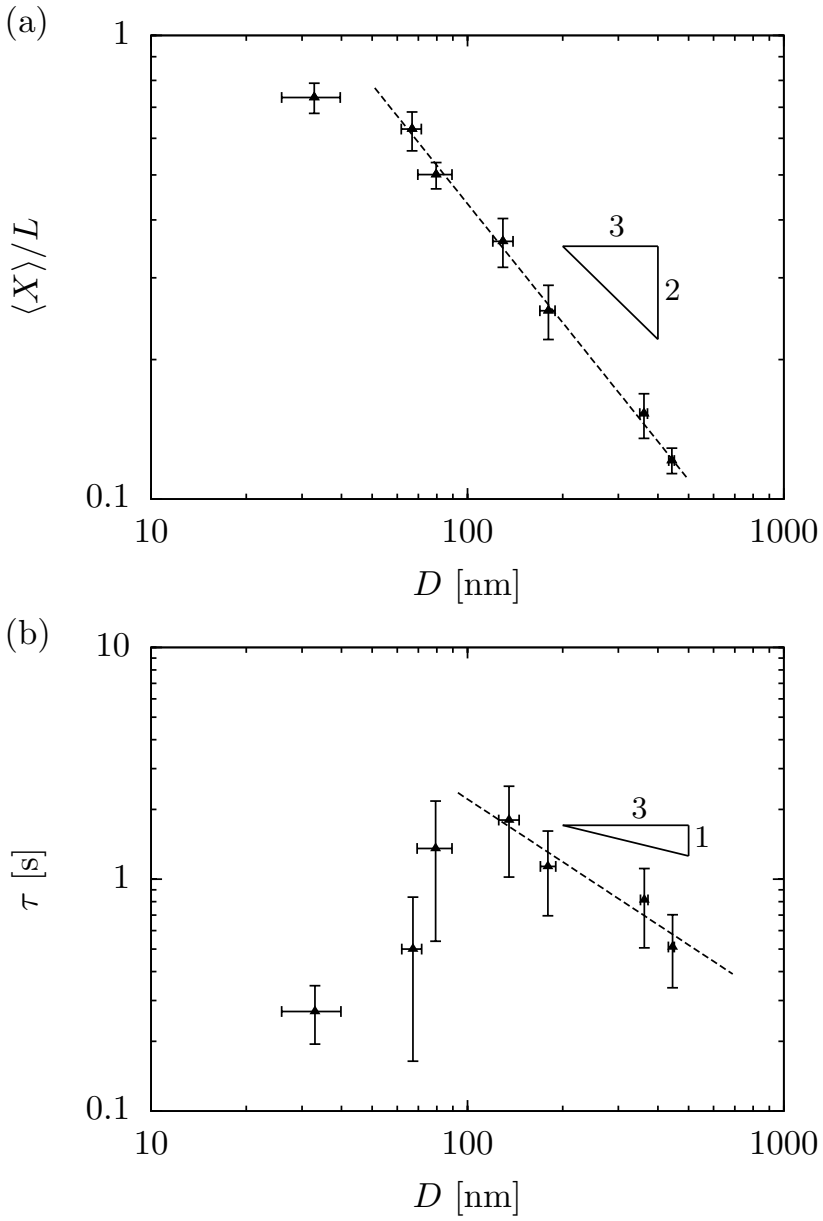

Figure 2.6: Reisner data for the extension (a) and relaxation time (b) of DNA in a nanochannel [60]. For the extension (a), De Gennes' theory predicts a scaling of $-2/3$, but the slope of the line of best fit is -0.85. The predicted slope for the relaxation time (b) is $-1/3$, but the slope of the line of best fit is -0.9.

Whether or not the Odijk regime can be observed for DNA also remains somewhat of a mystery. Unlike the de Gennes regime, it has proven difficult to fabricate channels and load DNA into sufficiently small to observe the Odijk regime. The smallest channels in Figure 2.6 are approximately 35 nanometers, which is clearly less than persistence length, but not sufficient to satisfy $D \ll l_p$. Nevertheless, while we are focusing here exclusively on DNA, it is worth mentioning stunning videomicroscopy images of actin filaments [134] that conclusively demonstrate the existence of an Odijk regime. Actin has a very large persistence length, so one can obtain a strongly confined chain using micron-sized channels. Further experimental evidence comes from measurements of a DNA-protein complex with a very large persistence length [135]. In the latter study, experimental data are in excellent agreement with extensions predicted by Eq. 2.44.

The resolution of these apparent discrepancies is a work in progress by the field, and constitutes several chapters in this dissertation (See Refs. 30, 71, 136 and Chapters 5 and 6). All of the proposed interpretations rest on the idea that intermediate regimes exist in between the Odijk and de Gennes regimes. The most developed of these interpretations is shown in Figure 2.7, and is based on Monte Carlo simulation results by Wang *et. al.* for DNA in nanochannels [71].

As the channel size decreases past the edge of the de Gennes regime, we enter a so-called "extended de Gennes" regime where the chain now looks like a series of anisotropic blobs [137]. This change in blob shape is induced when the contour length per blob reaches the thermal blob length; the blobs then deform anisotropically instead of decreasing the contour length per blob below this point. In this regime, the extension is the same as in the de Gennes regime [30, 71, 136, 138],

$$\langle X/L \rangle \approx (wl_p)^{1/3} D^{-2/3} \quad \text{(extended de Gennes)} \tag{2.45}$$

hence the moniker "extended" de Gennes. However, the predicted free energy in the extended de Gennes regime differs from the de Gennes regime. As such, we might expect to observe different dynamics in these two regimes even if the scaling law for the extension remains fixed [71].

Between the Odijk and extended de Gennes regime, there exists yet another regime where it appears that the scaling for the chain extension follows the behavior [71, 139,

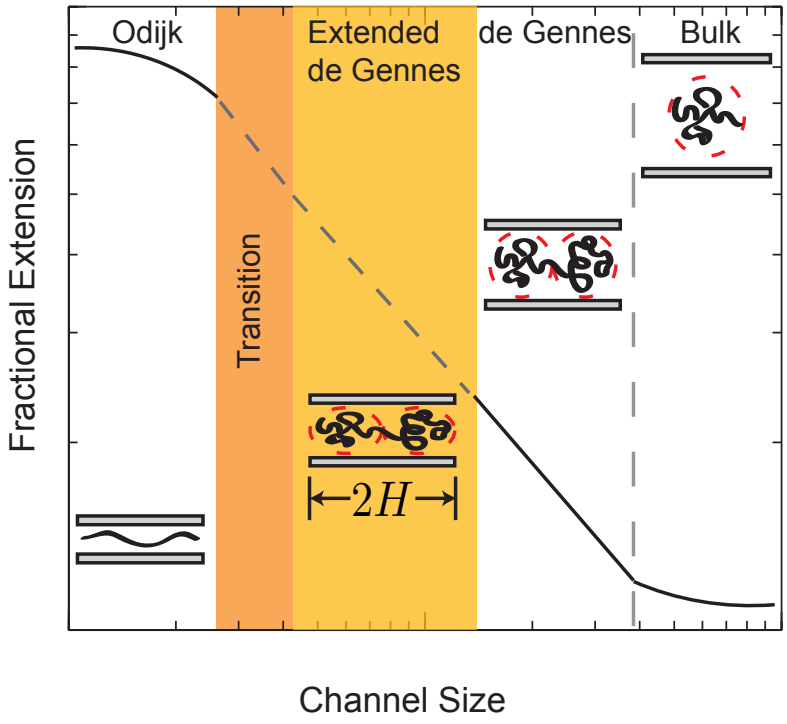

Figure 2.7: Qualitative sketch of the regimes of extension for a DNA chain as a function of the channel size D. The schematics show the qualitative models for the configurations of a confined chain. Figure reproduced from Ref. [31].

140]

$$\langle X/L \rangle \sim D^{-1} \quad \text{(transition)} \tag{2.46}$$

The details surrounding this regime are not well understood, and we simply refer to it as a "transition" regime between the Odijk and extended de Gennes regime. Finally, we summarize the fractional extension scaling regimes as well as the currently understood regime limits for a "long" piece of DNA in a high-ionic strength buffer in 2.1.

From the analysis of the physics of confined DNA, it is clear that we would like to be in the Odijk regime in order to maximize stretching and obtain the most sensitive measure of genomic length. Stretching in the Odijk regime will also insure that we have no hairpin folds and that we are able to resolve site-specific probes with maximum accuracy [75]. If we are in the Odijk regime, theory and empirical evidence suggest that the maximum resolution of a single measure is on the order of 1 kilobase pair [57,

Table 2.1: Channel sizes for the regimes of confinement for DNA in a high ionic strength buffer.*

Regime	$\langle X/L \rangle$	Channel Size
Bulk	$L^{-2/5}$	$D > R_g$ **
De Gennes	$D^{-0.701}$	$531\,\text{nm} < D < R_g$
Extended de Gennes	$D^{-0.701}$	$100\,\text{nm} < D < 531\,\text{nm}$
Transition	D^{-1}	$50\,\text{nm} < D < 100\,\text{nm}$
Odijk	2.44	$D \ll 50\,\text{nm}$

**i.e.* 5× TBE buffer. At this ionic strength (165 mM) [66], $l_p = 50\,\text{nm}$ and $w = 4.7\,\text{nm}$; **For comparison: λ-DNA, $R_g \approx 1\mu\text{m}$.

75]. However, as previously discussed, this is not a trivial experimental task, and it appears that for the moment we must resign ourselves to working with DNA in weaker confinement. Indeed, from Figure 2.6 and Table 2.1 it appears that most, if not all, of the relevant experimental data are in intermediate regimes. As such, we spend Chapters 5, 6 and 7 exploring this region of the parameter space, with a goal of better understanding DNA as a confined wormlike polymer.

2.5 Nanoslit Confinement

The distinction made in the literature between the title nanoslit and nanochannel is hazy, and sometimes "nanochannel" is used for any slit or groove where one dimension is less than a micron in size. Indeed for physical reasons, it is impossible to construct a true unbounded slit, since any supported structure must have walls. However the geometry most often referred to as a nanoslit is a wide channel where the aspect ratio exceeds 10:1 (width, W, to height, H). While we accept this definition as the colloquial definition for organizational purposes, it is important to make a further technical distinction. For a polymer chain to be practically considered in slit confinement, we need the walls to be far enough apart in the direction normal to the ceiling, so that the polymer does not feel their effects. This is best illustrated by 2.8, where we see three different channels of varying width. True slit confinement is illustrated on the right, with the polymer

chain manifesting an in-plane radius of gyration, $R_{\|}$, much smaller than the channel width. Below some critical width, the slit width W becomes smaller than the in-plane radius of gyration, and the polymer should then be considered in a high-aspect ratio nanochannel, rather than a slit.

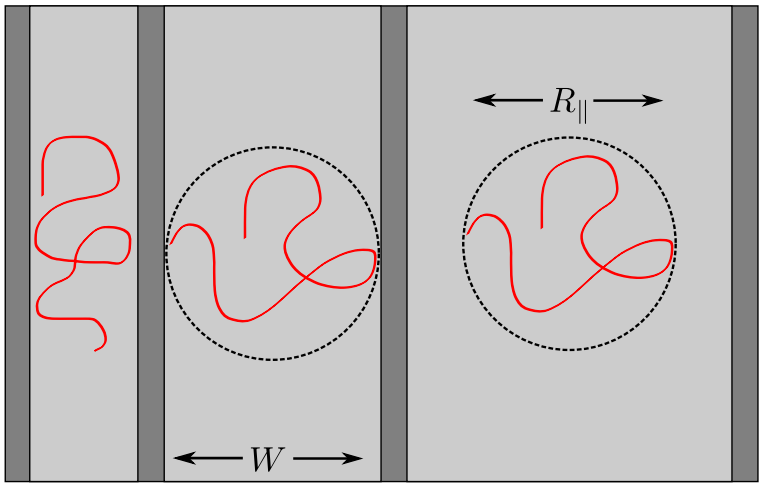

Figure 2.8: Polymer confinement in a nanochannel (left), high-aspect ratio nanochannel at the critical point where the slit width, W, equals the in plane radius of gyration (center), and slit with true quasi-2D confinement. Figure reproduced from Ref. 31.

The physics of wormlike chains in nanoslits parallels that in nanochannels, with some notable exceptions, and is schematized in Figure 2.9. For $H \gg R_g$, the chain is weakly confined and for $R_g \ll H$ compression blobs appear, giving rise to a de Gennes regime [136, 141, 142]. However the scaling of the de Gennes regime in slits is different than in channels. A Flory theory for a 2D walk of self-avoiding blobs is useful to show this is the case. In this theory, the free energy is given by

$$\frac{F}{k_B T} \sim \frac{X^2}{N_{\text{blob}} H^2} + N_{\text{blob}}^2 \frac{H^2}{X^2} \tag{2.47}$$

where N_{blob} is the number of blobs in the chain. Differentiating to find the minimum and using $H = L_{\text{blob}}^{3/5} (l_p w)^{1/5}$ to find the number of blobs gives

$$\frac{X}{L} \sim \left(\frac{w l_p}{H L} \right)^{1/4} \tag{2.48}$$

Importantly, the fractional extension in Eq. 2.48 does not scale linearly with L and actually decreases with increasing chain length ($X/L \sim L^{-1/4}$). This fact, which has been verified experimentally [143], implies that slits will not stretch as well as channels for a given minimum confinement dimension and is certainly a problem for genomic mapping applications. This is in contrast to nanochannels, which are critically scaled by the chain length ($X/L \sim L^0$). The loss in critical scaling is due to the fact that in slits, unlike channels, an increase in chain length decreases the local volume fraction of the chain [144].

By further decreasing the slit height, the amount of chain per blob decreases, until the chain can no longer form isometric blobs. As is the case in channels, intermediate regimes are proposed between the de Gennes and Odijk regimes. The leading candidate for the regime immediately adjacent to the de Gennes regime is an extended de Gennes regime, analogous to its counterpart in nanochannels [142, 145]. Here as the slit height decreases below the size of a thermal blob, excluded volume interactions cause the chain to swell axially forming "pancake blobs" with an in-plane diameter larger than the slit height. The extension in this regime is given by Eq. 2.48 and again we expect the free energy and dynamics to differ somewhat from the de Gennes regime.

As the slit height decreases further, we expect that blobs can no longer be formed and that a regime dominated by deflection segments will emerge [136, 142]. The nature of this transition was studied by a pair of conflicting experimental studies by Bonthuis *et. al.* [146] and Tang *et. al.* [147]. The former study suggested that the transition was sharp, and that the chain behavior quickly entered an Odijk regime. The latter study instead observed a broad transition over a large range of chain extensions, and did not observe an Odijk regime. After repeated study by both experiments [148, 149] and theory [142, 150, 151], it appears that the latter study has been confirmed and that the transition to an Odijk regime is broad and continuous.

Surprisingly, unlike the channel case, details surrounding the Odijk regime in slits remains a subject of some confusion in the literature. Articles by Burkhardt *et. al.* [131, 132] and Chen and Sullivan [141] study ideal wormlike chains in slit confinement, giving detailed theories and precise estimates of the relevant prefactors. By contrast, Doyle *et. al.* [142, 147] and Cifra [140] have looked at more realistic models, which incorporate excluded volume, and they obtain more modest results. Indeed with excluded volume is

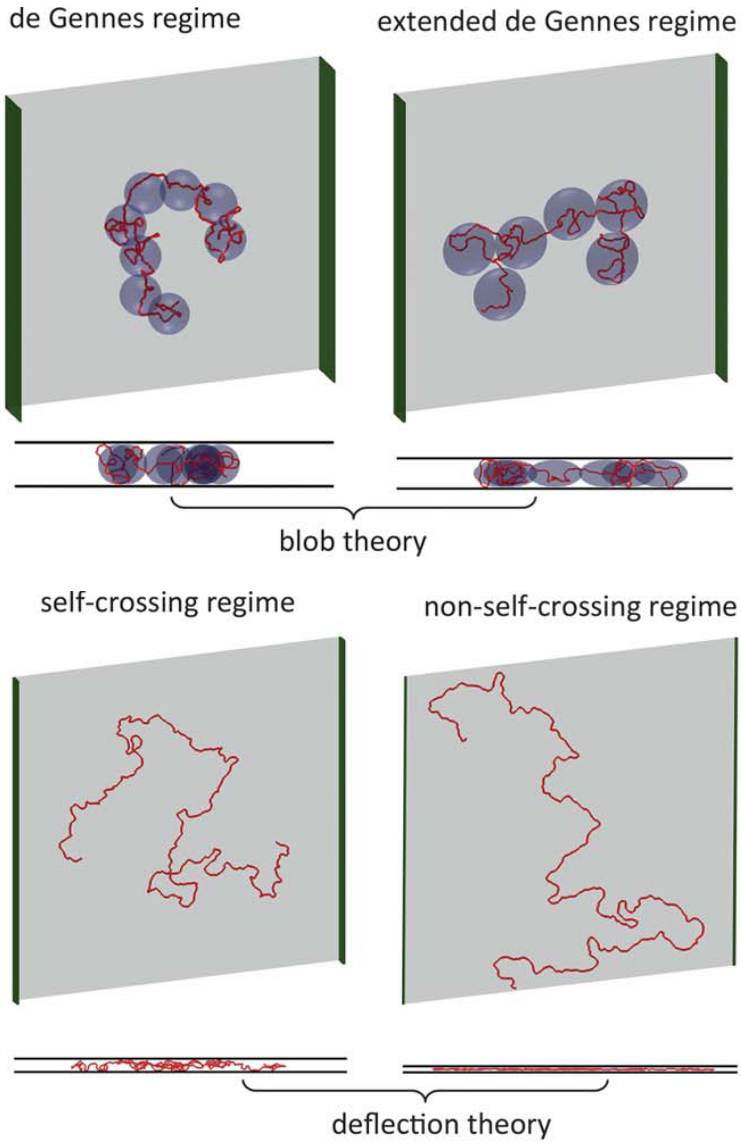

Figure 2.9: Four regimes postulated by Dai *et. al.* [142] for the extension of wormlike chains in slits. The regimes span from large slit heights in the upper left corner (de Gennes) to small slit heights in the lower right corner (Odijk). Figure reproduced from Ref. 142

included, one study suggests that there are two Odijk regimes for slits [142], and another suggests that there are none [140]. Other work by Jo *et. al.* [51] derives results for "slits" that are better characterized as high aspect ratio nanochannels per our discussion of Figure 2.8 above. Their expression for the fractional extension [51]

$$X/L \approx 1 - 0.085 \left[(H/l_p)^{2/3} + (W/l_p)^{2/3} \right] \tag{2.49}$$

contains both a slit height H and width W, and is proportional to L^0, suggesting that it is indeed scaling like a channel. Because of the confusion on the topic, we look at the Odijk regime in slits more closely in Chapter 8.

Finally, we note that while nanoslits are not very useful for stretching DNA, they are considerably simpler to fabricate than nanochannels and have been widely used for studying polymer dynamics in confinement. Both polymer diffusion [66, 152–155] and relaxation [146, 147, 156–159] are non-trivially influenced by the presence of slit walls. Indeed, the relaxation of DNA from a stretched state is observed to qualitatively change when slit walls are introduced [156]. In confinement, two separate time scales for relaxation emerge: one where the molecule relaxes into blobs the size of the slit height and one where the blobs relax in a quasi-2D random walk. In the bulk, only one time scale dominates the relaxation.

Diffusive behavior also shows an interesting difference in behavior between the bulk and confinement. Here, there is change in the screening of hydrodynamic interactions due the presence of the wall [152–154]. This results in measurements of the scaling of the diffusivity which disagree with the predictions of Brochard and de Gennes in Eq. 2.41 This "partial-draining" behavior is related to the results in Chapter 6 on the diffusivity of wormlike chains in channels. Indeed, the work in Chapter 6 inspired a collaboration with the Doyle group, where we rationalize the partial draining behavior in terms of an effect of the large monomer anisotropy of DNA [155].

Chapter 3

Monte Carlo Simulations

3.1 Introduction

In Chapter 2, we outlined a discrete wormlike chain model and stated our desire to compute equilibrium and dynamic properties relevant for genomic mapping. The governing equations for this model include the wormlike chain Hamiltonian in Eq. 2.3, the equilibrium probability density in Eq. 2.1 and the Fokker-Planck equation in Eq. 2.18. No closed form solution can be found for either the static or dynamic probability density for the free-solution Hamiltonian, let alone when wall boundary conditions are included. Therefore, we must resort to numerical methods to calculate the properties we desire.

Before we begin a numerical calculation, we face a fundamental choice between a Monte Carlo (MC) algorithm and a Brownian Dynamics (BD) method. MC algorithms sample the equilibrium probability density according to Eq. 2.1, whereas BD is able to calculate fully fluctuating chain dynamics according to Eq. 2.18. The tradeoff for the extra information resulting from BD is a dramatic increase in computational expense for equilibrium properties. Indeed, since MC algorithms are not constrained to reproduce physical dynamics, they are able to sample the phase space more efficiently. As such, MC methods are able to make calculations with $O(10^3)$ beads in a single chain on modern computers, whereas BD calculations (with hydrodynamics) are mostly limited to $O(10^2)$ beads. In the case of the Pruned-Enriched Rosenbluth method, which we introduce in Section 3.2.2, we are able to efficiently sample chains up to $O(10^5)$ beads.

Because of the extra expense incurred by Brownian Dynamics and because most of

the quantities which we desire are equilibrium properties, we restrict our scope to Monte Carlo methods only. As such, Section 3.2 explains the basics of Monte Carlo integration and the concept of importance sampling. Section 3.2.1 then introduces Metropolis Monte Carlo (MMC) calculations and introduces specific techniques for MMC simulations in confinement. Section 3.2.2 discusses chain-growth Monte Carlo algorithms and describes our implementation of the pruned-enriched Rosenbluth method (PERM). Finally, Section 3.3 expands on Section 2.3 and discusses methods for calculating the Kirkwood diffusion in confinement using Monte Carlo simulations. In addition, this section highlights the computational fluid dynamics (CFD) method we used to obtain the Green's function for a point force confined in a channel.

3.2 Monte Carlo Integration

We would like to be able to calculate a variety of equilibrium properties for our numerical model, including the ensemble averaged span, $\langle X \rangle$, rms end-to-end distance, $\langle R \rangle$, Helmholtz Free energy, $\langle F \rangle$, and entropy, $\langle S \rangle$. These properties fall into two categories. The first are mechanical properties, like the span and end-to-end distance, which depend explicitly on the phase space coordinates, and which must be determined from integrals like [160]

$$\langle X \rangle = \int X(\mathbf{r})\psi(\mathbf{r})d\{\mathbf{r}\} \tag{3.1}$$

where

$$\psi(\mathbf{r}) = \frac{1}{Z}\exp[-\beta U(\mathbf{r})] \tag{3.2}$$

is the conformational probability density function and

$$Z = \int \exp[-\beta U(\mathbf{r})]d\{\mathbf{r}\} \tag{3.3}$$

is the configurational partition function. The second are thermal properties, like the Helmholtz free energy and the entropy, which depend directly on the volume of phase space and therefore on the magnitude of the partition function [160].

To determine these properties, we need a numerical method to evaluate integrals like Eq. 3.1 ad Eq. 3.3. Because the dimensionality of the phase space is large, simple numerical quadratures (*e.g.* Simpson's rule) are intractable due to the need to finely discretize the phase space with such methods. Instead, we consider a class of integration

techniques based on the use of random variables and statistics, called Monte Carlo integration.

In Monte Carlo integration, we seek to obtain the value of the quadrature

$$I = \int_a^b y(x)dx \tag{3.4}$$

on the interval $x \in [a, b]$. Here, we focus on the one-dimensional case first, to avoid obfuscating the basic principles of the method with a complex, multi-dimensional integral like Eq. 3.1. In order to evaluate the the integral in Eq. 3.4, we pick uniformly random values of x along the interval, evaluate the function $y(x)$ and find the mean $\langle y(x) \rangle$. The mean-value theorem of basic calculus then relates the mean to the quantity of interest

$$I = (b - a) \langle y(x) \rangle \tag{3.5}$$

A 1D example of this technique is shown in Figure 3.1(a). Here we specify the function $y(x) = \exp(-x^2)$ and a lower ($a = 0$) and upper limit ($b = 5$) for the integral. Making 100 function evaluations from values of x uniformly distributed along the axis yields a value of $I_{100} = 0.553$. This is a poor estimate of the value of I, since its true value is $\sqrt{\pi/4}\,\mathrm{erf}(5) \approx 0.886$.

With a more careful selection of the points of x, we can improve the estimate of I without increasing the number of functional evaluations [161]. This is accomplished by choosing a biasing distribution $p(x)$ for sampling the function $y(x)$ in regions which are more important in the integral. The concept of picking a biasing distribution is fundamental to Monte Carlo integration and is labeled importance sampling [161]. In this scenario, the biased mean becomes,

$$\langle y(x) \rangle = \int_a^b p(x)y(x)dx \tag{3.6}$$

and the quadrature in Eq. 3.4 is evaluated by taking the average,

$$I = \left\langle \frac{y(x)}{p(x)} \right\rangle \tag{3.7}$$

to remove the bias.

Figure 3.1(b) shows the integration of our toy problem with a biasing distribution of $p(x) = c\exp(-x)$, where c is a normalization constant so the integral of $p(x)$ is equal

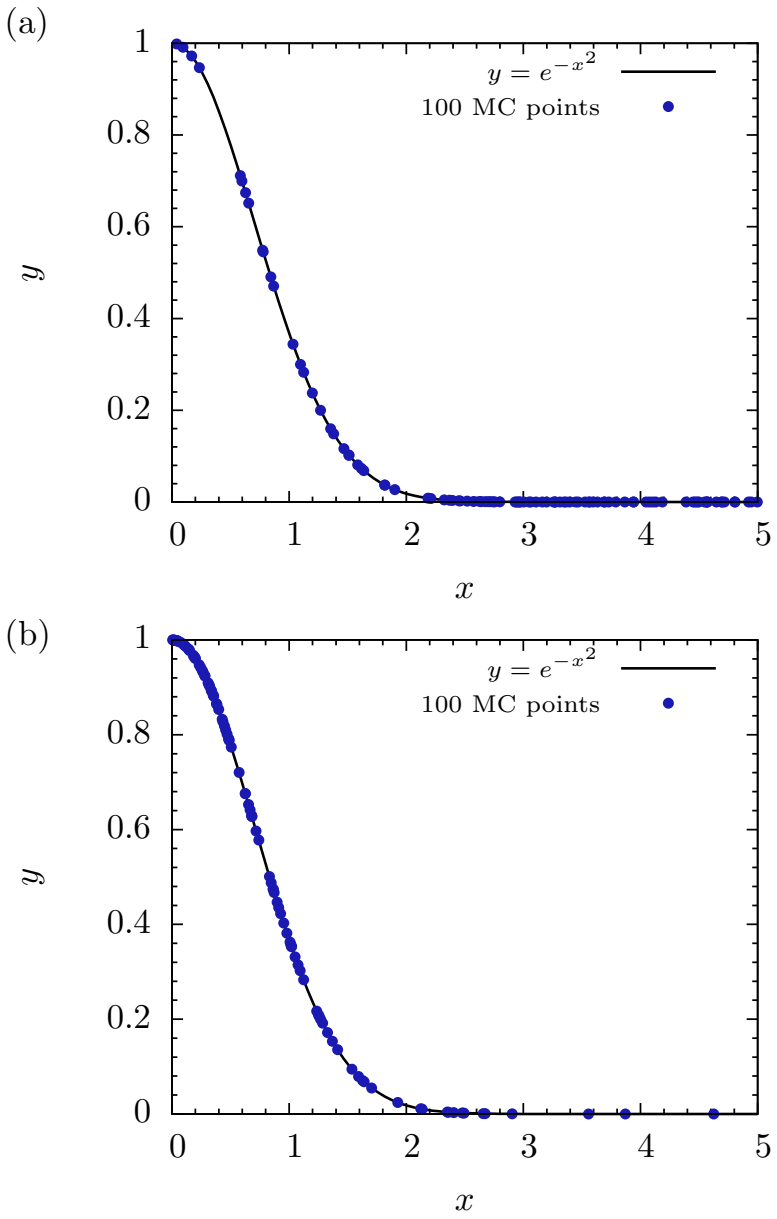

Figure 3.1: Monte Carlo integration of the function $y(x) = \exp(-x^2)$ using 100 function evaluations both (a) without importance sampling and (b) with importance sampling. The integration in panel (a) gives a value of 0.553; panel (b) gives a value of 0.913. The true value is $\sqrt{\pi/4}\,\mathrm{erf}(5) \approx 0.886$.

to one. Notice that the MC points are placed much more heavily upon the region where $y(x)$ is larger. Using this biasing distribution, the estimate of I_{100} is 0.913 — much closer to the value of 0.886.

As Eq. 3.7 implies, picking an efficient biasing distribution is the central challenge for Monte Carlo algorithms. Indeed, without good sampling Monte Carlo methods suffer the same fate as deterministic quadrature methods, and uselessly waste function evaluations in vacuous regions of the phase space. This is especially important in statistical mechanics, since the phase space distribution function, ψ, is generally very narrow, meaning much of the phase space makes negligible contributions to the partition integrals.

In principle, the best case scenario is where $y(x)/p(x)$ is a constant, and every function evaluation is given the perfect weighting [161]. However, it is not trivial to pick the optimal $p(x)$ *a priori*, especially in statistical mechanics where the biasing function is highly multi-dimensional. As such it is useful to divide the different approaches to importance sampling into two broad categories. The first class of techniques are so-called biased sampling techniques, where one uses a weighting procedure to directly or indirectly pick $p(x)$. The Pruned-Enriched Rosenbluth method falls into this category of methods and we defer a discussion of these methods to Section 3.2.2. The other class of methods eliminates the need to directly pick a biasing technique, and instead uses Markov chain theory to perform importance sampling [161].

3.2.1 Metropolis Algorithm

To examine Markov chain Monte Carlo methods, we return to our notation for the fully multi-dimensional integrals in Eq. 3.1 and Eq. 3.3. In Metropolis Monte Carlo (MMC), a Markov chain is constructed consisting of a set of states $\{\mathbf{r}\}$, which correspond to the set of possible chain conformations and transition probabilities, Π, which govern the motion between each state. A schematic showing a representation of the Markov chain and transition probabilities is shown in Figure 3.2. Basic results from the theory of stochastic processes states that if the transition probabilities Π are not a function of time (time homogeneity) and if every state is accessible (irreducibility), then we are guaranteed to reach a stationary distribution $\psi(\mathbf{r})$ at long times, regardless of the initial state [162].

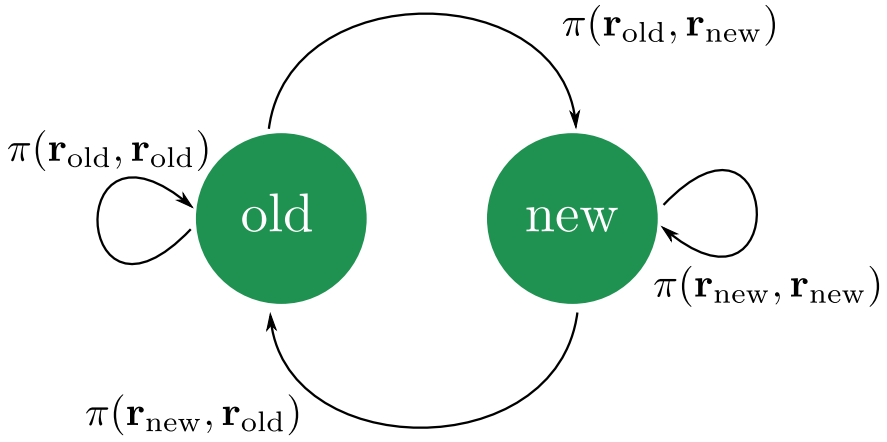

Figure 3.2: Markov chain method for obtaining a stationary distribution, ψ, for importance sampling. The stationary distribution is determined by the transition probabilities, π, between old and new states of the configurations of the chain.

As indicated by our notation, we would like to make our stationary distribution exactly equal to the equilibrium probability density function ψ. This is accomplished by setting the transition probabilities. A sufficient condition for obtaining the Boltzmann distribution in given by the detailed balance condition [161],

$$\psi(\mathbf{r}_{\text{old}})\Pi(\mathbf{r}_{\text{old}}, \mathbf{r}_{\text{new}}) = \psi(\mathbf{r}_{\text{new}})\Pi(\mathbf{r}_{\text{new}}, \mathbf{r}_{\text{old}}) \tag{3.8}$$

Substituting Eq. 3.2 into this expression and rearranging gives

$$\frac{\Pi(\mathbf{r}_{\text{old}}, \mathbf{r}_{\text{new}})}{\Pi(\mathbf{r}_{\text{new}}, \mathbf{r}_{\text{old}})} = \exp(-\beta\Delta U) \tag{3.9}$$

where $\Delta U = U(\mathbf{r}_{\text{new}}) - U(\mathbf{r}_{\text{old}})$. Since this expression is given as a ratio, we have some freedom in picking the transition probabilities. Standard practice is to use the Metropolis criteria [161]

$$\Pi(\mathbf{r}_{\text{old}}, \mathbf{r}_{\text{new}}) = \begin{cases} \exp(-\beta\Delta U), & \Delta U > 0 \\ 1, & \Delta U \leq 0 \end{cases} \tag{3.10}$$

which gives the Metropolis Monte Carlo technique its name.

We can interpret this rather esoteric derivation with the following practical description of the algorithm. The method begins with an initial configuration $\mathbf{r}_0$ upon which

a trial move is performed. The trial move is some kind of change to the configuration, such as a bead displacement, that alters the potential of the chain. If the change in potential ΔU is less than or equal to zero, the move is accepted. If it is greater than zero, the move is accepted with a probability given by the Boltzmann factor. The process is then repeated on the resulting configuration many times, until the sequence of states reaches equilibrium, and the configurations are visited with a frequency according to Eq. 3.2. Once the process has reached the equilibrium stationary distribution, ensemble averages from Eq. 3.1 can be computed by a simple mean.

Importantly, one never obtains a direct description of the partition function in the Metropolis MC algorithm — only the ratio of probabilities between chain conformations is necessary to perform trial moves. As a consequence, thermal properties are incalculable in the traditional implementations, though specialized algorithms have been developed to address this [160, 161]. The biasing methods discussed in the next section do not have this problem and thus the lack of the inherent ability to obtain thermal properties remains one of the weak points of MMC methods.

The choice of trial moves is key to the efficiency of a Metropolis Monte Carlo simulation. Because one is not limited to physical moves, it is also a potential source for creativity and innovation. The most efficient moves maintain a high probability of acceptance and also allow one to move a large distance in phase space. Normally there are trade-offs between these properties, since large moves in phase space tend to generate high-energy configurations that are unlikely to be accepted. Once the moves are implemented, the efficiency of the overall algorithm can be measured by the autocorrelation function of a measured property (*e.g.* the radius of gyration). Shorter relaxation times mean that the method generates independent configurations in fewer MC steps, indicating a more efficient algorithm.

The most common moves for off-lattice, rigid-body models such as our discrete WLC model are the reptation [163, 164], pivot [165] and crankshaft [164, 166], moves shown in Figure 3.3. A simple bead displacement move is not allowed for rigid-body models, since one must maintain a fixed distance between adjacent beads. In the reptation move, shown on the left, one of the ends of the chain is chosen at random. The bead from that end is removed from the configuration, and a new bead is placed at the other end of the chain. For apparent reasons this move is also sometimes called a slithering snake

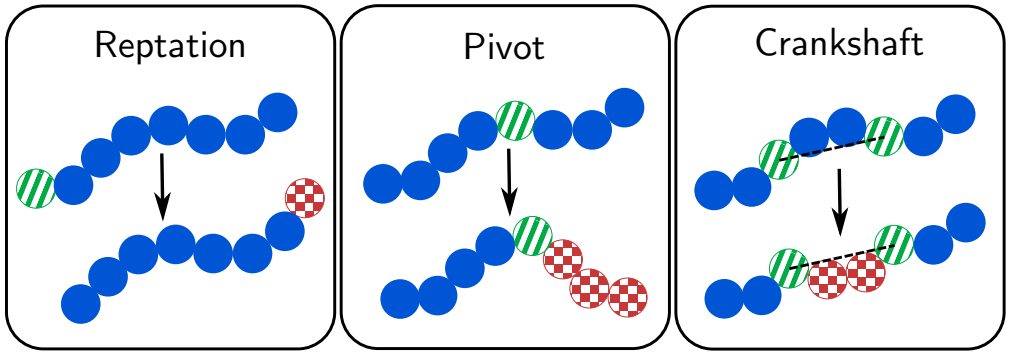

Figure 3.3: Monte Carlo moves for the wormlike chain model. "Selected" beads are marked with stripes and beads whose position changes during the move are marked with a checkerboard pattern. Other beads are solid filled.

move, which is synonymous with the word reptation.

The move shown in the middle panel of Figure 3.3 is the pivot move. In the pivot move a single bead in the chain chosen at random in addition to a random axis and two angles of rotation. One end of the chain is then randomly chosen to "pivot" about the axis by the random angles. The pivot algorithm can generate a large move in phase space and is often very efficient for chains in free solution. However, the pivot move has a particularly poor probability of acceptance in confinement, because it generates many conformations which overlap with walls.

Finally, the crankshaft move is shown in the right panel of Figure 3.3. In the crankshaft move, one chooses two beads at random in the chain, which are used to define an axis of rotation. All beads between the randomly chosen beads are rotated a random angle about this axis.

3.2.2 Chain Growth Methods

In contrast to Metropolis Monte Carlo, biased sampling methods attempt to pick a state from scratch, instead of repeatedly applying trial moves to a given configuration until an equilibrium is reached. For polymer chains, this amounts to an attempt to "grow" a chain, monomer by monomer until the desired chain length is reached. As such, biased sampling techniques are sometimes labeled static chain-growth methods. They are also called static Monte Carlo methods, since there are no Markov chain dynamics like there

are in the Metropolis algorithm. Notably, there is no limit to our ability to evaluate structural or thermal integrals in biased sampling methods. Furthermore, since all states are picked independently, we can begin counting statistics from the first state, instead of waiting for a Markov chain of states to relax to a stationary distribution.

The main challenge for static MC methods is the choice of an appropriate and efficient bias. Similar to our 1D example in Section 3.2, we would like to sample states in a region of the phase space that makes significant contributions to the integral. Since we have an enormous phase space from which to sample, the choice of the bias is critical to our evaluation of the integral. If the bias is too broad, then the algorithm is inefficient, and we waste time calculating configurations that make a negligible contribution to the relevant integral. A more subtle error happens if the bias is too narrow. In this case, the integral will appear to converge, but the lack of sufficient sampling will produce a biased estimate of the property of interest.

To see the importance of biasing for chain growth methods, consider the evaluation of the mean span and free energy of a 2D self-avoiding walk (SAW) on a square lattice as shown in Figure 3.4. The most straightforward algorithm to grow the chain is termed simple sampling. In simple sampling, a monomer is placed at a point on the lattice and monomers are repeatedly added, in random directions to the end of the chain. This is repeated until the chain reaches the desired contour length, resulting in a random walk of N steps.

To calculate the mean span and the free energy, we need to evaluate Eq. 3.1 and Eq. 3.3. First, we evaluate the integral of the partition function, which is a straightforward sum over M generated configurations

$$Z = \sum_{i}^{M} \exp(-\beta U_i) \tag{3.11}$$

noting that for a SAW, the Boltzmann factor is either zero or one. Since this is the case Z is less than 1 and the free energy $F = -\ln Z$ is a positive quantity. The mean span is also straightforward,

$$\langle X \rangle = 1/Z \sum_{i}^{M} X_i \exp(-\beta U_i) \tag{3.12}$$

However, just like our unbiased 1D example in Section 3.2, simple sampling does not sample the phase space of a SAW very well. In fact, the probability of generating a

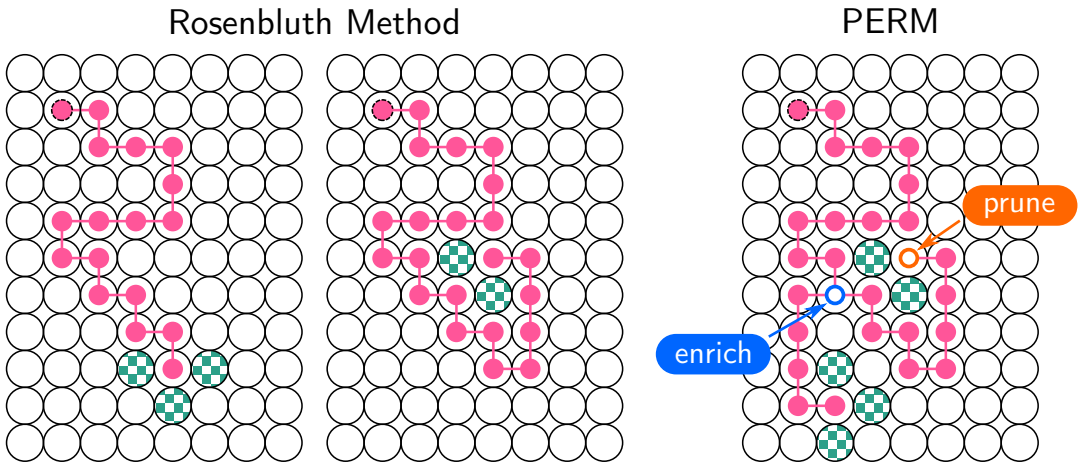

Figure 3.4: Self-avoiding walk on a 2D square lattice; steps are marked by filled circles. The left shows an example of Rosenbluth sampling where the next monomer is limited to unoccupied, adjacent lattice sites (checkered circles). The middle shows the failure mode of the Rosenbluth method, where the chain growth enters a trap. PERM overcomes traps by adding enrichment steps (hollow circles), as shown on the right. When the weight of the chain drops, branches are also pruned (hollow circles).

2D self-avoiding walk with a non-zero Boltzmann factor exponentially decreases as the chain length increases [167]. This problem is called attrition, and it is so severe for simple sampling that chains beyond a few dozen steps cannot be sampled efficiently.

The Rosenbluth method helps with attrition by adding a local bias to the chain growth step [168]. In the Rosenbluth method, one takes into account the directly adjacent occupied lattice sites, instead of growing the "next" monomer in a random direction. This is shown on the left in Figure 3.4, where there are only three valid choices for the next monomer position, instead of four. This added local bias must be taken into account in the integrals, and to do so, we introduce the Rosenbluth weight, W. The Rosenbluth weight is the inverse of the biasing probability, and is approximately equal to the Boltzmann factor of the chain [161]. Consequently, the partition sum becomes

$$Z = \sum_{i}^{M} W_i \tag{3.13}$$

and the mean span is given by

$$\langle X \rangle = 1/Z \sum_{i}^{M} X_i W_i \tag{3.14}$$

The addition of the local bias enables the simulation of chains nearing several hundred steps, instead of the several dozen possible with simple sampling. Nevertheless, even with the added bias, the Rosenbluth method still produces chains with an exponential attrition rate [167]. This is due to so-called "traps", which are visualized in the the middle panel in Figure 3.4. Here chain growth is able to proceed for a single step only, and then no remaining sites will be available. Such traps are a result of the myopic focus of the Rosenbluth method on the "next" step and its failure to anticipate the result of a series of steps. One method to avoid traps is a generalization of the Rosenbluth method, where one enumerates k steps into the future before placing the next monomer. This method is called the scanning method [169] and is useful for a small k, but still gives exponential attrition due to more difficult traps.

Instead of scanning future possible pathways, one may use enrichment [170] to get around traps, as shown in the right panel of Figure 3.4. In enrichment, one periodically generates branches, giving an exponential increase in chains to make up for the chains lost due to attrition. The "optimal" enrichment rate is therefore equal to the attrition rate, which is highly dependent on the specifics of the model and the geometry. Consequently, it becomes tedious to manually perform enrichments and some automated process is useful.

The pruned-enriched Rosenbluth method (PERM), invented by Grassberger in the late 1990's, is an extraordinarily useful example of such a process [171]. In PERM, instead of growing a single chain, a tree of chains (called a tour) is grown. As the tour grows, the Rosenbluth weight of the chain W_n is tracked and used as a metric to control enrichment and pruning. When the weight gets large, indicating a region of phase space where attrition is low, the tour branches. When the weight drops, attrition is likely and the algorithm prunes the tour. In this way, the tour prunes and enriches to counteract the traps that naturally occur from Rosenbluth sampling.

Modern implementations of PERM are very good at overcoming the attrition problem and can generate SAWs upwards of 10^4 steps [172]. This is demonstrated in Figure 3.5, which shows the probability of the survival of an N_{step} chain for simple sampling,

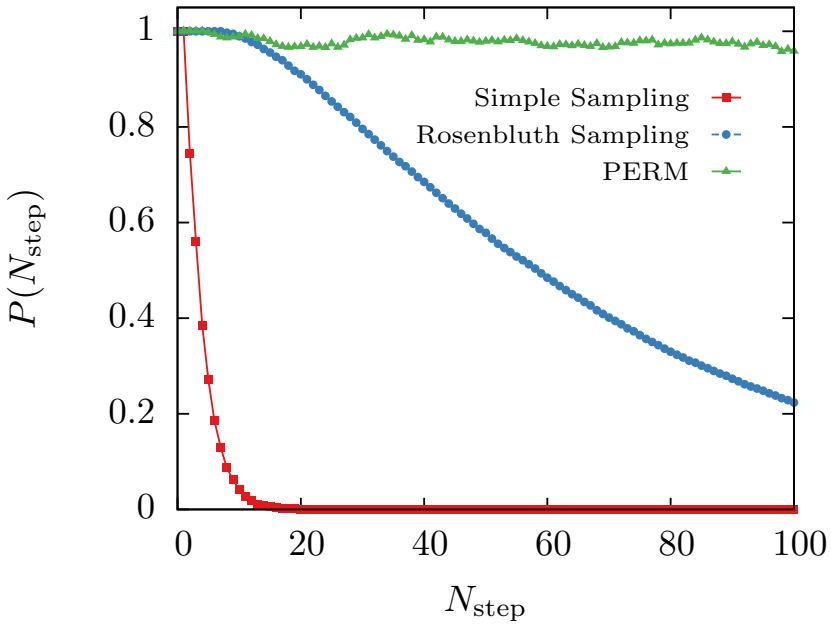

Figure 3.5: Probability of survival of an N_{step} chain for simple sampling, Rosenbluth sampling and PERM. Both simple sampling and Rosenbluth sampling decay exponentially, although Rosenbluth does markedly better than simple sampling. PERM gives an approximately flat probability with respect to step size as originally observed by Prellberg and Krawczyk [173].

Rosenbluth sampling and PERM. Remarkably, as Prellberg and Krawczyk noted [173], PERM gives a uniform distribution in chain lengths instead of the decaying exponentially like Rosenbluth sampling. This property is possible because the tours in PERM are correlated, so not all of the samples are independent. Thus the failure mode of PERM is different than with simple sampling and Rosenbluth; systems with large attrition rates generate very correlated tours that lead to poor statistics.

Another peculiarity of PERM is the need to estimate the weights as a function of chain length, W_n. Indeed, because the method uses the weights to control pruning and enriching, an accurate estimate of the weights is essential to an efficient algorithms. Indeed, if an inaccurate value of W_n is used, the method can become unstable and tours can "explode" with unrestricted enrichments or "die" due to over-pruning. Most often W_n is unknown at the beginning and must be obtained by some type of "bootstrapping"

method. In our experience, this must be done sequentially in chain length. Short chains simulations are run "blind" to obtain W_n for low n, and then an extrapolation is performed to estimate W_n at higher n. This extrapolation is used to in a larger simulation, and the process is repeated until the desired chain length is reached.

We used an off-lattice version of PERM in our simulations described in Chapters 4, 5 and 8. While the idea for off-lattice PERM was developed in the original paper by Grassberger [171], PERM has been almost exclusively used for lattice models. Unlike lattice versions, the enumeration of the weights in our version of the code is approximate, since the partition function is infinite for continuous space. However, the ratio Z/Z_0 where Z_0 is some reference partition function is not infinite. In our simulations, we have found it useful to use an ideal wormlike chain as the reference state for these simulations. Further details regarding this reference state and how it pertains to simulations of confined chains are found in the relevant chapters.

Finally, there are several other considerations to take into account when attempting to implement an efficient PERM algorithm. Notably, most data analysis must be done on the fly, which means all metrics must be calculated in $\mathrm{O}(n)$ time in a single pass (where n is the total chain length). Details for how this can be done are found in Appendix A. In addition, since tours are independent that can be run in a (mostly) embarrassingly parallel fashion on a supercomputing cluster. This greatly lowers the wall clock time needed to obtain good statistics.

3.3 Dynamic Properties from Monte Carlo

To close this chapter, we turn our attention from Monte Carlo integration to the calculation of the Kirkwood diffusivity using Monte Carlo data. In Chapter 2, we defined the Kirkwood diffusion tensor using the double-sum formula,

$$\boldsymbol{\mathcal{D}}_{\mathrm{cm}} = \frac{1}{N_b^2} \sum_{i=1}^{N_b} \sum_{j=1}^{N_b} \langle \boldsymbol{\mathcal{D}}_{ij} \rangle \tag{3.15}$$

for a chain with N_b beads of size a. In Eq. 3.15 the bead diffusivity tensor is given by

$$\boldsymbol{\mathcal{D}}_{ij} = k_B T \left(\frac{\delta_{ij}}{\zeta} \mathbf{I} + \boldsymbol{\Omega}_{ij} \right) \tag{3.16}$$

where $\zeta = 6\pi\eta a$ is the bead friction coefficient and $\mathbf{\Omega}_{ij}$ is the hydrodynamic mobility tensor.

For a Metropolis Monte Carlo simulation in free solution, the application of Eq. 3.15 is straightforward. A MMC simulation is run, and M independent configurations are obtained. These configurations are chosen by calculating the autocorrelation of $\boldsymbol{\mathcal{D}}_{ij}$ between successive configurations and picking configurations separated by a number of Monte Carlo steps longer than the relaxation time. Using these configurations, Eq. 3.15 is evaluated using the Oseen–Burgers tensor

$$\mathbf{\Omega}_{ij} = (1-\delta_{ij})\mathbf{\Omega}_{ij}^{\mathrm{OB}}(\mathbf{r}_{ij}) \tag{3.17}$$

$$= \frac{1}{8\pi\eta r_{ij}}\left(\mathbf{I} + \frac{\mathbf{r}_{ij}\mathbf{r}_{ij}}{r_{ij}^2}\right) \text{ for } i \neq j \tag{3.18}$$

and then averaged over the M configurations.

Unfortunately, for general geometries the hydrodynamic tensor (*i.e.* Green's function) does not have a closed-form analytical solution akin to $\mathbf{\Omega}_{ij}^{\mathrm{OB}}$. To see this, consider Stokes equations [114],

$$-\nabla P(\mathbf{r}) + \eta\nabla^2\mathbf{u}(\mathbf{r}) = -\sum_{i=1}^{N_b}\mathbf{f}_i\delta(\mathbf{r}-\mathbf{r}_i) \tag{3.19}$$

which, when combined with the continuity equation,

$$\nabla\cdot\mathbf{u}(\mathbf{r}) = 0 \tag{3.20}$$

describes the pressure and velocity fields of the solvent interacting with a system of N_b point particles exerting a set of forces, $\mathbf{f}_i$. For a single particle in an unbounded domain, the solution to Eq. 3.19 gives the Oseen–Burgers tensor appearing in Eq. 3.17. This is sufficient to solve Eq. 3.19 for many particles, since the system of equations (Eq. 3.19 and Eq. 3.20) is linear and the principle of superposition applies.

Confining geometries enter the problem through no-slip and no-penetration boundary conditions. Using the method of images, Blake [174] found the Green's function for Eq. 3.19 and Eq. 3.20 near a single planar wall with a no-slip condition. Similarly, Liron and Mochon [175] found an infinite series solution for the velocity and pressure fields of a fluid perturbed by a point force between two plates. However, to the best of

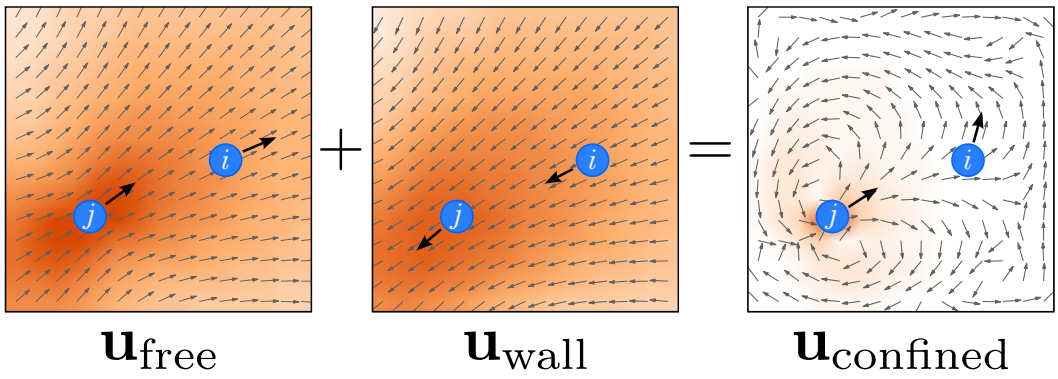

Figure 3.6: Method for calculating the hydrodynamic Green's function in confinement. The free solution velocity field $\mathbf{u}_{\text{free}}$ (left) is added to a wall velocity field $\mathbf{u}_{\text{wall}}$ (middle) giving the confined velocity field $\mathbf{u}_{\text{confined}}$ (right). While the free solution velocity has a known analytical solution, the wall velocity does not and must be obtained numerically.

our knowledge, no analytical solution can be found for the velocity field due to a point force in a channel or circular tube. In fact, even the series solution of Liron and Mochon requires numerical evaluation to be practically useful in this case.

Accordingly, we resort to numerical methods to obtain a solution. Unfortunately, a direct computational fluid dynamics (CFD) calculation is impossible due to the point forces in Eq. 3.19. However, we can get around this using a method pioneered by Jendrejack *et. al.* [176] shown in Figure 3.6. Following Jendrejack, we divide the two difficult parts of the calculation, (i) the point forces and (ii) the no-slip boundary conditions, into two separate systems of equations. Since the Stokes equation and the continuity equation are linear, this is acceptable if the sum of the two new systems of equations are equivalent to Eq. 3.19 and Eq. 3.20 with $\mathbf{u} = 0$ on all of the boundaries.

To deal with the point forces, the first system is assigned to be the "free solution" case given by Eq. 3.19 and Eq. 3.20 as written above with an unbounded domain. As previously stated, the Green's function for this system of equations is the Oseen–Burgers tensor, and we denote the velocity from this system with the term $\mathbf{u}_{\text{free}}(\mathbf{r})$. The second system of equations is given by

$$-\nabla P_{\text{wall}}(\mathbf{r}) + \eta \nabla^2 \mathbf{u}_{\text{wall}}(\mathbf{r}) = 0 \tag{3.21}$$

where we define

$$\mathbf{u}_{\text{wall}}(\mathbf{r}) = \mathbf{u}(\mathbf{r}) - \mathbf{u}_{\text{free}}(\mathbf{r}) \tag{3.22}$$

To account for the no-slip boundary condition, we recall that in the original system, $\mathbf{u} = 0$ everywhere on the boundary so the boundary conditions for the "wall" system are

$$\mathbf{u}_{\text{wall}}(\mathbf{r}) = -\mathbf{u}_{\text{free}}(\mathbf{r}) \text{ (boundary only)} \tag{3.23}$$

The solution to Eq. 3.21 and Eq. 3.23 is not known analytically, but because it has no point forces, it can be solved by conventional fluid dynamics techniques.

The resulting hydrodynamic tensor for confined systems is given by

$$\mathbf{\Omega}_{ij} = (1 - \delta_{ij})\mathbf{\Omega}_{ij}^{\text{OB}} + \mathbf{\Omega}_{ij}^{\text{wall}} \tag{3.24}$$

where the wall term is obtained by repeated CFD calculations. Repeated calculations need to be made because (i) we desire to know the hydrodynamic tensor as a function of space and (ii) because only velocities can be obtained from CFD calculations. To do this, a point in space $\mathbf{r}$ is chosen as well as a force vector $f\mathbf{e}_\alpha$ along some unit vector in an orthogonal coordinate system (*i.e.* $\alpha = x, y, z$). From this, the boundary velocities are calculated from $\mathbf{\Omega}_{ij}^{OB}$ and the simulation is performed. The resulting wall velocities can be used to compute 3 of the 9 components of the hydrodynamic tensor according to

$$\begin{pmatrix} \Omega_{xx}^{\text{wall}}(\mathbf{r}) \\ \Omega_{yx}^{\text{wall}}(\mathbf{r}) \\ \Omega_{zx}^{\text{wall}}(\mathbf{r}) \end{pmatrix} = \frac{1}{f_x}\mathbf{u}_{\text{wall}}(\mathbf{r}) \tag{3.25}$$

This process is then repeated for the other two unit vectors and across the entire spatial range of interest.

We developed a custom code to perform the CFD simulations because it was difficult to incorporate the appropriate boundary conditions into existing packages and because significant cost savings could be obtained for repeated calculations. Accordingly, we used a second-order, finite-difference numerical method to discretize Eq. 3.21 and Eq. 3.20 with the boundary conditions in Eq. 3.23. A staggered grid was used to avoid velocity "checkerboarding" and to get around known complications for the

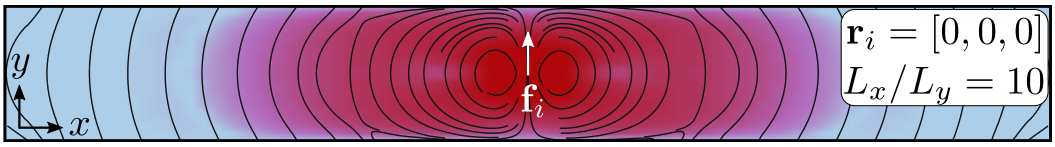

Figure 3.7: Velocity from a point force from our custom CFD code on a projected surface in a nanochannel. The figure shows streamlines (solid lines) and the logarithm of the magnitude of the projection of the velocity vector (unspecified units, dark shaded regions are four orders of magnitude larger than light shaded regions) onto the $z = 0$ plane. The force is placed at the origin and directed along the positive y-axis, normal to the wall.

pressure boundary conditions for incompressible systems [177]. We solved the resulting linear system of equations using GMRES with an incomplete LU decomposition preconditioner. Importantly, because finite-difference methods do not guarantee mass conservation, we found it necessary to define "flux-averaged" boundary conditions. To do so, the mass flux J over a portion of the boundary surface with area A was obtained using a two-dimensional Gaussian quadrature of the free solution velocity. The boundary condition for $\mathbf{u}_{\text{wall}}$ was then defined as $-J/A$, to satisfy both Eq. 3.23 and mass conservation.

Sample results from this code, similar to those used in Chapters 6 and 7 are shown in Figure 3.7. To approximate an infinitely long channel, we used a large aspect ratio (10), which was sufficient to see the velocity drop several orders of magnitude. As such, we assumed translational symmetry in the channel axis and confined our point forces $\mathbf{f}_i$ to the $x = 0$ plane. Finally, not that in all calculations in Chapters 6 and 7, the values for the Green's function were stored on a grid and off-grid values were determined by five-dimensional linear interpolation [178].

Chapter 4

Free Solution DNA as a Wormlike Chain

This chapter is based on the publication:
D. R. Tree, A. Muralidhar, P. S. Doyle and K. D. Dorfman
Macromolecules 46 (2013), 8369–8382

4.1 Introduction

Double-stranded DNA (dsDNA) has long stood as a unique polymer due to its role in biology and biochemistry. In addition, thanks to modern techniques in molecular biology and soft-matter physics, monodisperse samples of dsDNA can be prepared with an extraordinarily large range of molecular weights, which can in turn be visualized and controlled at the single-molecule level. Accordingly, dsDNA has assumed the role of a "model polymer" and has been extensively studied. Despite its widespread use, accumulating evidence suggests that dsDNA is *not* a good model polymer for investigating universal polymer properties, and a version of the more flexible single-stranded DNA (ssDNA) with limited base pair interactions has been proposed as an alternative [109, 179]. In this paper, we examine the length-dependent properties of both single-stranded and double-stranded DNA in order to further evaluate their fitness as model polymers.

In order to do so, we first ask why any specific polymer would be an appropriate general model in the first place? The answer is given by the aptly named concept

of *universality*, which was well explained by de Gennes [77]. Universality implies that at sufficiently large length (and time) scales, all dilute solutions of self-avoiding polymers in a good solvent exhibit equivalent behavior, regardless of disparate underlying chemical structures [69, 72, 77]. Therefore at large enough contour lengths *all* polymers are "model" polymers, because all polymers behave similarly (*e.g.* entropic elasticity and self-avoidance). This is certainly the sense in which dsDNA, ssDNA or any other polymer is meant as a model polymer.

Due to the specific chemical structure of dsDNA, its behavior is well described by the wormlike chain model [64], and at short enough length scales (near the persistence length) dsDNA is often described as semiflexible. Accordingly, it is sometimes repeated that "dsDNA is not a good model of flexible molecules, because it is semiflexible". However, this statement can lead to confusion due to an unfortunate ambiguity surrounding the word "semiflexible" that often arises in the literature. Per our definition of universality, this statement is correct if the term semiflexible is meant to denote a polymer with a contour length near its persistence length. Indeed, universality provides no basis for comparing any short-chain polymer to either another short-chain polymer of different chemistry or the general behavior of all polymers. Using this terminology, a flexible chain is thus any chain that shows universal behavior. However, if the term semiflexible is used as a synonym for the class of polymers that are well described by the Kratky-Porod model (*i.e.* wormlike chains), then this statement is incorrect since very long wormlike polymers indeed show universal behavior. The statement is all the more misleading because it implies that dsDNA is always semiflexible, which is true if semiflexible is synonymous with wormlike, but false if semiflexible means short. In this paper, we shall use the term "semiflexible chain" to denote a member of the class of wormlike chains.

Regardless of notational convention, the principle of universality immediately suggests a way to assess the theoretical appropriateness of a proposed model polymer. Namely, is the polymer long enough such that chemically specific behavior disappears? This of course completely neglects the bedrock question that made DNA the model polymer of choice: Is the model polymer experimentally convenient to use? For a polymer to serve as both a correct and practically useful model polymer, these two questions

must be answered affirmatively. In this work, we purposefully omit any normative statements about experimental convenience, and instead compute the contour length where the chemically specific behavior of dsDNA disappears. In this way, we seek to quantify how both static (*e.g.* radius of gyration) and dynamic (*e.g.* diffusion coefficient) properties of dsDNA approach universal behavior. Along the way, we find it instructive to compare to the properties of a model of ssDNA as well.

There is a lack of consensus in the literature regarding the appropriate length at which one can consider dsDNA to be a flexible chain. For instance, some authors have claimed that even a very long molecule like λ-DNA is "ideal", being too short and stiff to experience excluded volume interactions [109, 180]. However other studies suggest that excluded volume interactions indeed have an impact at similar contour lengths [95, 181, 182]. Further confounding the issue, the oft-cited measurements of the diffusion coefficient of concatamers of λ-DNA by Smith et al. [89] suggested that dsDNA had already reached a universal limit. However, the work by Smith et al. is at odds with recent theoretical work on the draining behavior of wormlike polymer coils [183] and work on DNA in confinement [155] which suggests that the molecular weight required to reach the universal limit for dynamic properties is even larger than the weight required for static properties.

Adding to the confusion, intercalating dyes, which make dsDNA so convenient to use in fluorescence microscopy experiments, have an undeniable impact on the chain chemistry of dsDNA — extending the molecular contour length by 20-30% [184]. Even so, the most basic molecular properties of dyed dsDNA (*i.e.* the persistence length) remain difficult to accurately measure and are therefore controversial [185, 186]. And while we previously stated that universality implies that chain chemistry has no qualitative impact on the regime of universal behavior, a change in the persistence length or effective width can alter both the molecular weight of the transition as well as the limiting value of a specific molecular property (*i.e.* radius of gyration).

In order to assess the transition of dsDNA from short-chain to universal behavior, we adopt a numerical approach and compute static and near-equilibrium dynamic properties of both ssDNA and dsDNA as a function of molecular weight using a Monte Carlo algorithm. Specifically, we employ the powerful Pruned-Enriched Rosenbluth Method (PERM) which allows us to capture an enormous range of molecular weights of dsDNA

— from short oligonucleotides to near chromosomal lengths. While the application to DNA is unique, the numerical techniques we employ are not, and several excellent resources exist for the interested reader [172, 187, 188]. With the range and precision afforded by PERM, we are able to make specific quantitative predictions of measurable properties of very long dsDNA molecules and subsequently provide insight into the transition to long-chain, flexible behavior.

4.2 Model and Methods

4.2.1 DWLC Model

The discrete wormlike chain model (DWLC) described in Chapter 2 (see also Refs. [71, 189, 190]) is a coarse-grained polymer model, which in contrast to bead-spring models [191], is capable of capturing sub-persistence length behavior. As a key feature, the DWLC model is able to reproduce properties of both the freely jointed chain (FJC) and the continuous wormlike chain (CWLC), which makes the DWLC versatile enough to model both single- and double-stranded DNA. Double-stranded DNA has been modeled analytically as a CWLC [192], but since a numerical model is necessarily discrete, the DWLC model is appropriate when using small discretization lengths. By contrast, models with both discreteness and bending stiffness have been used for ssDNA, making the DWLC model an ideal choice [189, 193–195]. We note however that to use such a simple model for ssDNA, we must neglect specific base-pair or base-stacking interactions. Neglecting such interactions also means dismissing many important properties of ssDNA, but we hypothesize that this model will describe some important non-interacting sequences of ssDNA [179, 195, 196] or ssDNA in denaturing conditions. In order to proceed with a description of the DWLC model, we defer a more rigorous justification to Sec. 4.3.2, where we present our parameterization of the model to experimental measurements.

As introduced in Chapter 2, the model is defined as a series of N inextensible bonds of length a with a bending potential [71, 189, 190]

$$\beta U_{\text{bend}} = \kappa \sum_{j=1}^{N-1} (1 - \cos\theta_j) \tag{4.1}$$

between each pair of bonds. Here κ is the bending constant, β is the inverse temperature $(k_B T)^{-1}$ and θ_j is the angle formed between adjacent bonds j and $j+1$. With this definition, the contour length of the chain is given by $L = aN$. Note that our implementation does not incorporate bond extensibility, which can be important for modeling DNA under large tensile forces [189, 193, 197]. In practice, this is done by replacing the inextensible rods with a finitely-extensible bond potential.

Due to the simplicity of Eq. 4.1, the equilibrium probability density function for a bond angle can be written in closed form, which is useful for chain-growth simulations (see online supporting information). From this, one can obtain a relationship between the the bending constant, κ, and the Kuhn length, b [198, 199]

$$\frac{b}{a} = \frac{\kappa - 1 + \kappa \coth \kappa}{\kappa + 1 - \kappa \coth \kappa} \tag{4.2}$$

When $\kappa \gg 1$, this reduces to the familiar expression for a CWLC, $b/a = 2\kappa - 1$. When $\kappa \to 0$, Eq. 4.2 reduces to $b = a$, since the DWLC becomes an FJC in the limit of no bending potential. In referring to the chain flexibility, we often find it convenient to describe polymer flexibility by the persistence length, l_p, which is related to the Kuhn length, $l_p \equiv b/2$.

In addition to incorporating flexibility, space-filling chains require an excluded volume potential. To add excluded volume, $N+1$ spherical beads are introduced at the bond joints and a hard bead repulsion is defined at the diameter w by the potential

$$\beta U_{\mathrm{EV}}(r_{ij}) = \begin{cases} \infty, & |r_{ij}| \leq w \\ 0, & |r_{ij}| > w \end{cases} \tag{4.3}$$

where $|r_{ij}|$ is the positive distance between bead centers at i and j. The choice of hard beads over a finite potential increases program efficiency and simplicity and gives an athermal excluded volume model.

Eq. 4.3 suggests that the excluded volume potential U_{EV} is independent of the bond length, a. However, the choice of bond length does indeed affect the excluded volume behavior of the chain. When the bead radius is small compared to the bond length, $w < a$, unphysical chain crossing can occur, and if $w > a$, adjacent excluded volume beads may "overlap". In practice, w is set to be greater than or equal to a, since bead overlap is simple to overcome, but chain crossing is not. To prevent bead overlap in

a model with a substantial bending penalty at the bead length scale, one can simply redefine Eq. 4.3 to apply when $j > i+k$ where k is an arbitrary positive integer. (In our case we set $k = 2$.) The constant k defines a minimum length scale of self-interaction, a concept which is a commonly used in polymer field theories [69].

4.2.2 Numerical Method

To calculate equilibrium polymer properties with the DWLC model, we employ the Pruned-Enriched Rosenbluth method (PERM). PERM is a chain growth Monte Carlo algorithm that employs a dynamic bias to obtain importance sampling [171] and is distinct from Markov-chain (*i.e.* Metropolis) algorithms. PERM is an advanced method for long polymer chains and overcomes the well-known attrition problem that limited chain length in the Rosenbluth-Rosenbluth (RR) algorithm [168]. To do so, a tree of chains (called a tour) is grown according to a bias that is implemented by controlling the rates of pruning or enriching [170] of the branches of the tour.

In our off-lattice version of the algorithm, this is done as follows (see also Chapter 3). We initiate a chain at the origin and for the n^{th} chain growth step, we make K trial steps according to the probability distribution of the polymer bending potential (see online supporting information). Each trial step is assigned a Rosenbluth weight,

$$a_n^{(k)} = \exp(-\beta U_n^{(k)}) \tag{4.4}$$

where $U_n^{(k)}$ is the potential energy due to intrachain interactions. (In this case it is U_{EV}.) The weight of the growth step n is defined as

$$w_n = \sum_{k=1}^{K} a_n^{(k)} \tag{4.5}$$

and to make the step, one of the trial steps is randomly chosen according to the probability

$$p_n^{(k)} = \frac{a_n^{(k)}}{w_n} \tag{4.6}$$

The cumulative weight of the chain at step n is defined as

$$W_n = \prod_{i=0}^{n} w_i \tag{4.7}$$

which is an approximate count of the number of configurations using K trials. As the chain grows, W_n fluctuates and can become zero if a suitable self-avoiding chain cannot be found. To circumvent this, pruning and enrichment are used to bias the chain growth towards successful states. When W_n rises relative to its ensemble average $\langle W_n \rangle$, chain growth is deemed successful and the tour spawns branches (enrichment). Conversely, when $W_n / \langle W_n \rangle$ falls, chain growth is struggling, and the tour is pruned. This perpetual cutting and growing of the chain leads to a depth-first search type of diffusion along the chain contour length [171] and the method subsequently yields statistics as a function of molecular weight.

Our strategy for pruning and enriching follows a stochastic, parameterless version by Prellberg and Krawczyk [173], which we found to be simple and efficient. Unfortunately, the addition of Markovian anticipation [200] to our pruning and enriching scheme did not result in a significant speed-up, likely due to the large persistence length of the simulated chains.

Nevertheless, a significant reduction in the computational cost was achieved by a more mindful calculation of $\langle W_n \rangle$. Since W_n is generated during execution, $\langle W_n \rangle$ can be determined at run-time. However, the initial estimate is poor, which leads to slow execution (especially for large chains). We found, as expected [201], that $\log \langle W_n \rangle$ becomes linear in n for large n. This allowed us to run short chain "blind" [167] estimates of $\langle W_n \rangle$ and linearly extrapolate to large n, obviating the need to bootstrap our way to an estimate of $\langle W_n \rangle$ for large n. Importantly, this extrapolation does not bias the ensemble averages in any way, but simply increases the efficiency of the algorithm.

In addition, a careful enactment of $\mathrm{O}(n^2)$ procedures proved key for an efficient implementation of PERM. Since the chain growth requires $\mathrm{O}(n)$ operations, a naive implementation of an $\mathrm{O}(n^2)$ procedure at each step, n, yields calculations that scale like n^3. Efficient implementation is further hampered by the fact that recording each tour's configuration is prohibitively expensive (in both time and memory) and data analysis must be done on the fly. To circumvent the problem, properties such as the radius of gyration and diffusion coefficient were coded to iteratively update with each growth step, which kept the algorithm $\mathrm{O}(n^2)$ as desired. With additional scrutiny and a neighbor list, many property evaluations could be reduced to $\mathrm{O}(n)$ time (such as the radius of gyration and the form factor [187], see supporting information online), which

subsequently allowed for greater reductions in the required computational time.

In our implementation, we employed a master/slave parallel algorithm without Markovian anticipation on a DELL Linux cluster. We reach self-avoiding chains of up to 1×10^5 beads (for dsDNA), which is close to two orders of magnitude longer than our efforts with a conventional Metropolis algorithm [71], but still falls short of the exceptionally long chains in the newest implementations of the pivot algorithm [81, 202]. Recent work by others using PERM for semiflexible chains on a lattice have reached similar chain lengths [172]. Static properties were calculated with as many as 4×10^5 (dsDNA) and 5.3×10^5 (ssDNA) tours and dynamic properties were calculated with 10^5 (dsDNA) and 1.3×10^5 (ssDNA) tours. The batches of tours were divided into subsets in order to estimate the error (standard error), which is sufficiently small that the data shown in all figures is smaller than the given symbol size, unless otherwise depicted.

4.2.3 Properties

To assess the approach of dsDNA to universal values, we evaluate several static and dynamic properties. We are particularly interested in measures of the size of the chain that can be obtained experimentally. These include the radius of gyration

$$S = \left\langle \frac{1}{N+1} \sum_{i=1}^{N+1} (\mathbf{r}_i - \mathbf{r}_{cm})^2 \right\rangle^{1/2} \tag{4.8}$$

which can be measured by various scattering techniques, as well as the mean span [71]

$$X = \langle \max(\mathbf{x}) - \min(\mathbf{x}) \rangle \tag{4.9}$$

and the root-mean-square end-to-end distance

$$R = \left\langle (\mathbf{r}_{N+1} - \mathbf{r}_1)^2 \right\rangle^{1/2} \tag{4.10}$$

both of which can be measured by fluorescence microscopy. In these expressions $\mathbf{r}_i$ represents the (3×1) vector position of the i^{th} bead of the chain and $\mathbf{x}$ represents the the $(N + 1 \times 1)$ vector of all of the x positions in the chain. Note that, unless the polymer is confined, one typically obtains the diffusion coefficient in fluorescence microscopy, from which the end-to-end distance or radius of gyration is inferred.

The polymer form factor, commonly obtained by light scattering measurements, can also be obtained from simulation data using the relation [187]

$$P(q) = \left\langle \frac{1}{(N+1)^2} \sum_{i,j=1}^{N+1} \frac{\sin(qr_{ij})}{qr_{ij}} \right\rangle \tag{4.11}$$

where r_{ij} is the distance between beads i and j, and q denotes the magnitude of the scattering wave vector $\mathbf{q}$.

In addition to structural properties (*e.g.* radius of gyration) commonly obtained in all Monte Carlo methods, PERM can calculate thermal properties (*e.g.* entropy) as well. Observe that if the sum in Eq. 4.5 is replaced with an integral, the ensemble average of W_n (Eq. 4.7) corresponds to the definition of the configurational partition function [171, 173]. By performing repeated, stochastic chain-growth steps we are simultaneously sampling this integral (relative to an ideal chain standard state) similar to the Widom particle insertion method [161]. Accordingly, the excess free energy of a chain of length L due to interchain interactions is

$$\beta F = -\ln \left\langle \frac{W_N}{K^N} \right\rangle \tag{4.12}$$

It is also worth mentioning, that when hard potentials are employed $\langle W_N \rangle$ is simply a count of the configurations and the excess free energy reduces to the excess entropy.

In addition to the measures of static properties, it is possible to use PERM to estimate the near-equilibrium chain diffusivity by the so-called rigid-body approximation of the Kirkwood diffusivity [124, 203] (see also Chapter 3). We do so by giving the $N+1$ beads a hydrodynamic diameter d in an implicit continuum fluid, which — due to the small length scale — exhibits very small Reynolds number flows. Since the most important intrachain interactions come from beads that are far apart along the contour of the chain, we make the reasonable assumption that we can use a far-field approximation for the hydrodynamic interactions. The low-Reynolds number and far-field approximations yield an Oseen-Burgers tensor for the Green's function of the bead velocity. When this is combined with a first-order correction for the finite bead size, the chain mobility tensor becomes [122, 129, 176, 204]

$$\mathbf{\Omega} = \frac{1}{(N+1)^2} \sum_{i,j=1}^{N+1} \left[\left(\frac{\delta_{ij}}{3\pi\eta d} + \frac{1-\delta_{ij}}{8\pi\eta r_{ij}} \right) \mathbf{I} + \frac{\mathbf{r}_{ij}\mathbf{r}_{ij}}{{r_{ij}}^2} \right] \tag{4.13}$$

The Kirkwood diffusivity [114, 116] is given by the equilibrium ensemble average of the trace of the chain mobility tensor

$$D = \frac{k_B T}{3} \langle \mathrm{Tr}(\mathbf{\Omega}) \rangle \tag{4.14}$$

Equation 4.14 neglects the effects of dynamic fluctuations in the chain conformation and is thus an approximation (to within a few percent error [114, 205]) of the "true" near-equilibrium diffusion coefficient (which is given by the mean-square displacement or Green-Kubo relations) [121, 124]. Since Eq. 4.14 employs an ensemble average of a conformational property, it can be calculated from PERM, or any other Monte Carlo algorithm [203]. This enables us to calculate the diffusivity of a very long, semiflexible chain *with* excluded volume and hydrodynamic interactions, a feat that has proved extraordinarily difficult by analytical theories.

4.3 Results and Discussion

4.3.1 Review of Dilute Solution WLCs

To facilitate the discussion of the long-chain behavior of DNA, we briefly review some aspects of dilute solutions of wormlike chains. We focus on the case of a WLC in three-dimensions which is unperturbed by external forces and refer the otherwise interested reader to recent references on WLCs perturbed by forces [206, 207] and confined to planar surfaces [208]. As stated in Sec. 4.2.1, a continuous wormlike chain is characterized by a contour length L, a Kuhn length b, an effective chain width w and a hydrodynamic diameter d. By dimensional analysis, only three combinations of these parameters can be unique, giving us a three-dimensional phase space.

Neglecting chain dynamics for the moment, consider the equilibrium phase plane depicted in Fig. 4.1. The phase diagram divides the equilibrium behavior of WLCs into three universal regimes: rod, Gaussian coil and swollen coil [172, 209], which are conveniently explained by scaling arguments. Note that while the scaling theory outlined here provides a physical basis for the existence of the universal regimes, it is unable to address details regarding either transition regions or prefactors of a given property [69]. Indeed after reviewing the scaling theory, the object of much of the remaining discussion will be to compute and analyze the practical consequences of the prefactors

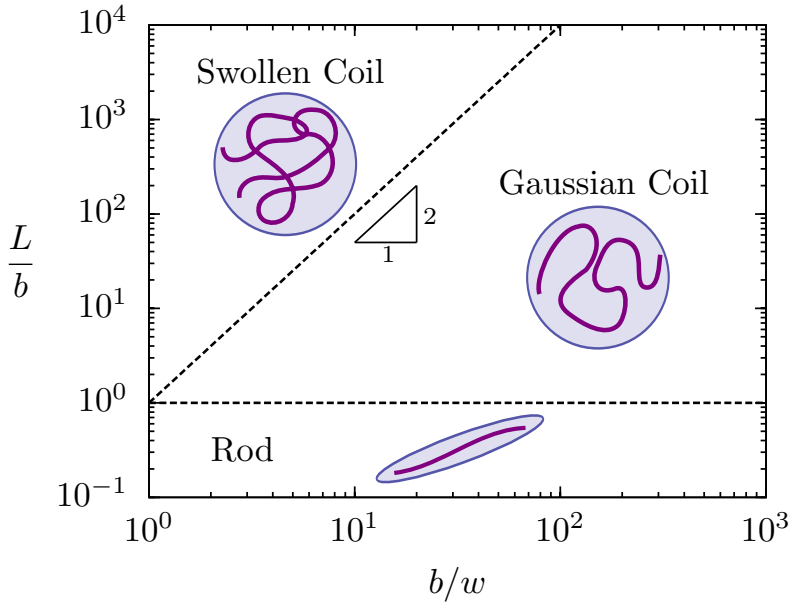

Figure 4.1: Summary of the classical scaling arguments for a real semiflexible chain in dilute suspension. Three regimes are predicted based on the interplay between the contour length (L), the chain stiffness (l_p) and the chain width (w). Very short chains ($L \ll l_p$) are rod-like and long "thin" chains are nearly Gaussian ($L \gg l_p$ and $L \ll l_\mathrm{T}$). Long chains ($L \gg l_\mathrm{T}$) are swollen coils.

and transition regions that scaling theory is unable to address. Since we have limited our scope to single- and double-stranded DNA, we direct the generally interested reader to recent work by Hsu et al. [172, 187] where a simpler lattice model is used to compute the prefactors and transition regions of a dilute WLC over a broad range of the parameter space.

Consider the case of a chain with a constant b/w, which would be a vertical trajectory in Fig. 4.1. When $L \ll l_p$, the chain is short and rigid like a rod and the size of the polymer — which we represent with the end-to-end distance R without a loss of generality — scales linearly with the contour length

$$R \sim L \tag{4.15}$$

When the chain is much longer than the persistence length $L \gg l_p$, the thermal fluctuations of the chain overwhelm the bending energy and the shape becomes a flexible coil. However, if the polymer is short enough, there are few intrachain interactions and

the molecule experiences negligible excluded volume interactions giving

$$R \sim (bL)^{1/2} \tag{4.16}$$

which is the familiar random walk scaling.

For any real (self-avoiding) chain, the magnitude of the total excluded volume interactions increases as the contour length increases. When the excluded volume energy is on the order of k_BT, a second transition occurs from a Gaussian to a swollen coil. The size of a swollen coil is given by the radius [72]

$$R \sim b\left(\frac{w}{b}\right)^{2\nu_F-1}\left(\frac{L}{b}\right)^{\nu_F} \tag{4.17}$$

where $\nu_F = 0.587597(7)$ [81] is the modern value of the Flory exponent.

Just as the rod-to-coil transition is characterized by the persistence length, the Gaussian-to-swollen coil transition is given by the contour length contained in a thermal blob

$$l_{\mathrm{T}} \equiv c\frac{b^3}{w^2} \tag{4.18}$$

with c given as a scaling constant. Normalizing Eq. 4.18 by the Kuhn length b reveals the dependence of l_{T} on the monomer anisotropy b/w, which is the ratio of the "stiffness" to the "thickness" of the chain [71, 109, 210]. Thus when $L \ll l_{\mathrm{T}}$, the chain is too stiff and thin to swell and the chain scales like Eq. 4.16, whereas when $L \gg l_{\mathrm{T}}$ the chain experiences sufficient excluded volume interactions to scale like Eq. 4.17.

An equivalent picture to the thermal blob (to within a constant factor) that often appears in the theoretical polymer physics literature [211, 212] is the excluded volume parameter [72, 210]

$$z \equiv \left(\frac{3}{2\pi}\right)^{3/2}\frac{w}{b}\left(\frac{L}{b}\right)^{1/2} \tag{4.19}$$

with the conventional prefactor. Here $z \approx 1$ signifies the transition point between Gaussian and swollen behavior.

4.3.2 Model Parameterization

Since we are interested in moving beyond scaling theory and making quantitative predictions of the properties of single-stranded and double-stranded DNA, we need to

parameterize the DWLC model to experimental data. In particular, we need values of the persistence length, l_p, (or equivalently the Kuhn length), the effective width, w, and the bond length, a, in order to specify equilibrium properties. To specify dynamic properties, we also need the hydrodynamic diameter, d, but we defer this discussion until Sec. 4.3.4.

As polyelectrolytes, the magnitude of the persistence length and the effective width of single- and double-stranded DNA depend on the ionic strength [66]. The effect of ionic strength, I, on the persistence length of double-stranded DNA has been examined by both experiments [63–65, 193, 213–215] and theory [216–218]. While some disagreements remain regarding the effects of electrostatics on l_p at very low values of the ionic strength, empirical results and theories generally agree for large values of I [66]. Perhaps due to lesser prominence or greater measurement difficulty, there seems to be little controversy surrounding the magnitude and ionic strength dependence of the effective width of dsDNA. More than three decades ago, Stigter used the calculation of the second virial coefficient of a stiff, charged rod to predict the width [66, 219], and it has been subsequently corroborated by DNA knotting experiments [220].

Figure 4.2 summarizes the ionic-strength dependence of dsDNA of both the Kuhn length (using an empirical relation from Dobrynin [66, 218]) and the effective width (using Stigter's theory [219]). As has been pointed out before [66], both b and w rise as the ionic strength decreases, making the chain stiffer. However, above 10 mM the Kuhn length changes much less quickly than the effective width, meaning that the monomer anisotropy ratio drops rapidly as the ionic strength decreases. Therefore as b/w falls with decreasing ionic strength, the Kuhn monomers of dsDNA become less anisotropic and the chain becomes more flexible in the long chain limit. The effect on the scaling behavior of finite length chains is non-trivial however, since there is a competition between decreasing the anisotropy of the Kuhn monomers and decreasing their number. For instance, near 10 mM the monomer anisotropy ratio for dsDNA falls below 10, but the number of Kuhn monomers in T4-DNA remains high ($\approx$ 400). However, by 0.1 mM the anisotropy ratio drops below 3, but the number of Kuhn lengths is reduced to approximately 100.

To simplify the model, we limit our scope to high ionic strengths where strong electrostatic screening marginalizes the effect of the electrostatics [66]. This assumption

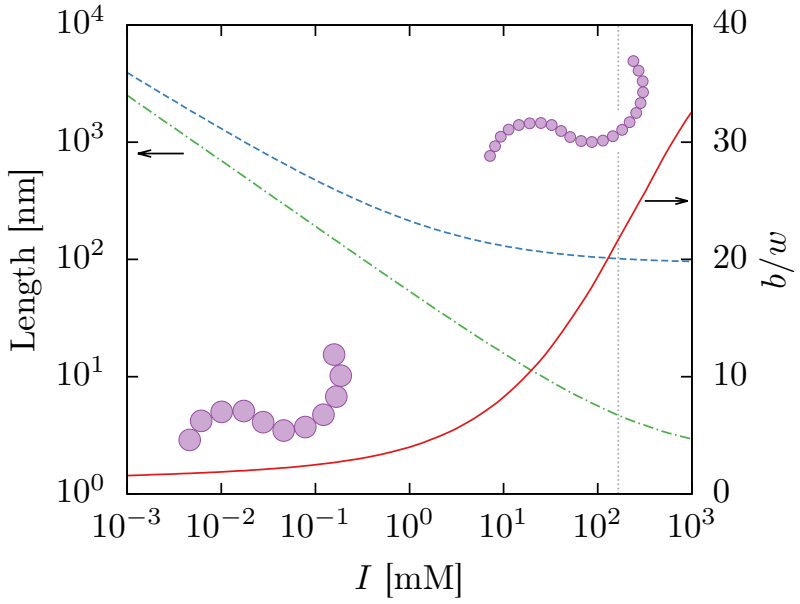

Figure 4.2: Ionic strength dependence of the Kuhn length (dashed line) and the effective width (dot-dashed line) and the monomer anisotropy (solid line) for dsDNA. The vertical dotted line indicates an ionic strength of 165 mM ($\approx$ 5$\times$ TBE) [66]. The schematic illustrates two chains with similar b but different w, demonstrating the decrease in monomer anisotropy as the effective width increases more rapidly than the Kuhn length.

allows us to neglect electrostatic potentials in our model and use constant values of b and w. When necessary, we assume an ionic strength of 165 mM, which corresponds to 5$\times$ TBE [66] and is marked with a gray vertical line in Figure 4.2. Assuming these buffer conditions gives $b = 106$ nm, which has become the consensus Kuhn length of dsDNA [64], and $w = 4.6$ nm, the value obtained from Stigter's theory [66, 71, 219].

Unlike dsDNA, even a simple measure of the persistence length of ssDNA in a high ionic strength buffer remains controversial. A survey of the recent literature reveals studies done by mechanical stretching [193, 221], fluorescence recovery after photobleaching [222], fluorescence resonance energy transfer [196, 197] and small-angle x-ray scattering [195, 196] that yield values of the bare persistence length (at infinite ionic strength) between 0.6 and 1.3 nm. It seems likely that base-base interactions are responsible for the disagreement and recent work on non-interacting ssDNA sequences [195, 196] gives a consistent value of 1.5 nm at the aforementioned ionic strength of 165 mM.

Assuming rod-like interactions for the effective width, which forms the foundation of Stigter's theory, appears to be inappropriate for ssDNA. Some experimental work suggests that for ssDNA, w is nearly independent of ionic strength [221] and that its value is approximately equal to the bare persistence length of the chain [195]. Accordingly, we adopt a value of 0.65 nm [195], which conveniently also appears to be the approximate rise of a single base of ssDNA [223].

Finally, because we have a discrete model, we must specify a bond length a. Since dsDNA is well-described by a continuous model, the choice of a is somewhat arbitrary so long as $a \ll b$, much like a time step in numerical integration schemes. In this case we choose $a = d$, which is the smallest length scale in the model. This is commonly called the touching-bead model [75, 122] and is also advantageous for the calculation of the diffusivity. In addition to the far-field approximation mentioned in Sec. 4.2.3, the DWLC estimation of the Kirkwood diffusivity also introduces discretization errors into the diffusivity; accurate hydrodynamic interactions require the collective action of many Stokeslets, which in turn requires a large number of beads. The touching-bead model provides adequate resolution of the chain to satisfy this condition and has the additional benefit of circumventing any artifacts in the hydrodynamics due to a variable bond length.

The choice of a for ssDNA is less clear than for dsDNA, since both continuous and discrete models have been used with some success for ssDNA [189, 193–195]. The Kuhn length (≈ 3 nm) provides the upper bound for a, and it is sufficiently small that the lower bound is given is given by the chemical monomer size (≈ 0.6 nm). As discussed in Sec. 4.2.1, a choice of $a > w$ is problematic, so for convenience we set $a = w$. For the reader's convenience, the model parameters are summarized in Table 4.1.

PERM calculations were performed with the parameters in Table 4.1 to verify the model. Results for the radius of gyration are shown in Fig. 4.3 where they are compared to experimental values of the radius of gyration of undyed dsDNA obtained by light or neutron scattering [82, 84, 85, 87, 88, 93, 95, 99, 103, 104, 106, 107]. There is excellent agreement between the experimental data and the PERM results. However, given that there are two degrees of freedom (l_p and w) to fit the experimental data, the good agreement between theory and experiment is expected.

Table 4.1: Parameters of the discrete wormlike chain for double- and single-stranded DNA in a high ionic strength buffer[*].

Parameter[**]	Symbol	ssDNA	dsDNA
Kuhn length	b	3.0	106
effective width	w	0.65[***]	4.6
hydrodynamic diameter, bond length	d, a	0.65	2.9

[*] $\approx 5\times$ TBE; [**] All of the parameters are lengths expressed in nm; [***] Note that while our model parameters are defined in some cases to sub-angstrom precision, this does not reflect the true experimental accuracy of these parameters.

Even with the excellent agreement, the effective width w remains a somewhat uncertain parameter. Eq. 4.17 predicts that $S \sim w^{0.175}$, demonstrating that the radius of gyration is not very sensitive to the effective width. This means that S is not particularly useful at evaluating the ability of the model to capture the correct strength of the excluded volume. Compounding this fact, it appears to be difficult to collect accurate data for the radius of gyration at large L.

Thus, despite our best efforts to pick accurate model parameters, our choice is certainly a possible point of contention. Indeed, our parameters give a monomer anisotropy $b/w \approx 23$ for dsDNA, whereas others have estimated an anisotropy as high as 66 [109]. Implicit in this disagreement, and further muddying the waters, is the role of intercalating dyes mentioned in the introduction. Both the persistence length and effective width are clearly affected by the presence of dyes [184], but consistent measurements of their effects have not been possible to date [185, 186]. Consequently, we have omitted data for dyed dsDNA from Fig. 4.3, since reported radius of gyration measurements are inferred from *diffusivity* measurements, which are not static measures. In fact, in Sec. 4.3.4, our results suggest that this introduces a source of systematic error since the chain has not yet reached the long-chain limit. Therefore, until more data regarding the effective width and a resolution of the dye-dependence of both l_p and w becomes available, the parameters for dyed dsDNA will remain somewhat uncertain.

Finally, in addition to the PERM calculations, we also found it useful to employ

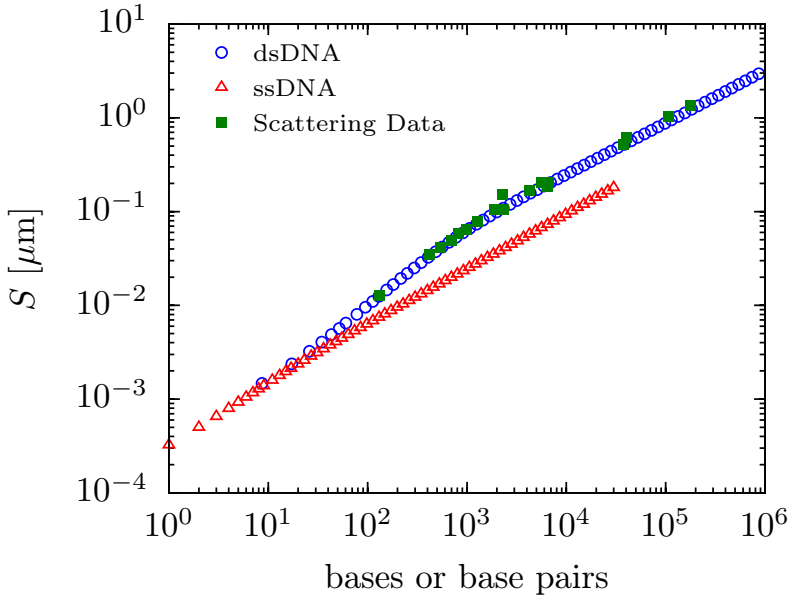

Figure 4.3: PERM data for the radius of gyration of ssDNA (open triangles) and dsDNA (open circles) compared to experimental data for dsDNA from light and neutron scattering (filled squares) [82, 84, 85, 87, 88, 93, 95, 99, 103, 104, 106, 107]. The experimental data were obtained from many different references, at varying ionic strength (all $\geq$ 100 mM), with varying information about the molecular weight. To obtain a consistent value, the molecular weight of dsDNA was assumed to obey the relation 660 Da = 0.34 nm = 1 bp. Single stranded DNA was assumed to follow 1 base = 0.65 nm.

renormalization group theory (RG) results by Chen and Noolandi [212] for the end-to-end distance and radius of gyration. In contrast to the Monte Carlo results, the RG theory gives R and S only, but the metrics are available as a function of molecular weight to practically unlimited contour lengths. Note that since the RG calculations employ a continuous model, the excluded volume strength in the RG theory must be re-parameterized to agree with the experimental and PERM data (see online supporting information).

4.3.3 Equilibrium Properties of DNA

Given that we have a parameterized model for single-stranded and double-stranded DNA, we are prepared to move beyond the insights of scaling theory and examine detailed quantitative calculations of dilute solution equilibrium properties. To begin, we examine the value of the apparent power-law exponent of the end-to-end distance

$$\nu \equiv \frac{\mathrm{d} \ln R}{\mathrm{d} \ln L} \tag{4.20}$$

as a function of molecular weight. While this can be done with PERM, the end-to-end distance is also available from the RG theory of Chen and Noolandi [212], which we parameterize to match the PERM data (see online supporting information). The RG theory is only available for only a few equilibrium properties, but it provides results over a much larger range of contour lengths than one can obtain with PERM. In addition, the RG theory has no sampling error and gives a much smoother value of ν.

Fig. 4.4 shows the RG theory results for the power-law exponent, ν, as a function of L/b. As a reference, results are shown for several different values of the monomer anisotropy, w/b, not just ssDNA and dsDNA. At first glance, nothing appears spectacular about the plot; it agrees very well with scaling theory. For instance, consider the curve corresponding to $w/b = 3.16 \times 10^{-3}$. When L/b small, the chain is rod-like and around $L/b = 1$, the exponent falls rapidly, approaching $\nu = 0.5$. As L/b increases, excluded volume gradually dominates and the exponent approaches 0.588. Furthermore, the dependence on w/b of the transition from Gaussian to swollen coils also agrees with scaling theory. For w/b near 1 (ssDNA), the chain shows effectively no Gaussian regime and the chain transitions from rod-like to swollen coil very quickly. Whereas, when w/b

is near 0 ($w/b = 3.16 \times 10^{-4}$) a large Gaussian regime appears and the transition to a swollen coil is delayed.

In addition, Fig. 4.4 supports the contemporary interpretation of ssDNA as a flexible chain. The scaling exponent transitions practically immediately to an excluded volume chain and by about 500 bases the exponent is 0.58 — within 1% of Flory scaling. (It should be noted that the RG calculations limit to a value of $\nu = 0.5886$ [212], slightly larger than the value of ν_F reported by Clisby [81].)

However, the asymptotic and continuous transitions between regimes in Fig. 4.4 are unknown from scaling theory, and these transitions have a major consequence on the implications of scaling theory for dsDNA. As the RG calculations show, double-stranded DNA is an intermediate case, with a monomer anisotropy $w/b \approx 0.04$ and is therefore neither thin nor stiff enough to exhibit a true Gaussian coil regime. In fact, the minimum exponent shown in Fig. 4.4 is $\nu = 0.535$ at 8.3 kilobase pairs (kbp), which is about halfway between 0.5 and ν_F. Additionally, the transition of dsDNA to a completely flexible coil is exceptionally broad [172]. At 48.5 kbp (λ-DNA), the exponent ($\nu = 0.546$) is only slightly higher than the minimum; by the time the chain reaches 1 megabase pair (Mbp), the exponent has reached 0.572, which is within 3% of ν_F.

Thus, the continuous and asymptotic nature of the transitions obfuscates the scaling theory picture of dsDNA. Accordingly, Fig. 4.4 provides an excellent explanation for the confusion surrounding the scaling of dsDNA. That is, for most practical purposes, kbp-length dsDNA is neither "ideal" or "real", but an intermediate case.

Given that the transitions are smooth, one would expect excluded volume effects to play a non-negligible role for dsDNA at intermediate contour lengths, well before the flexible chain limit is reached. This is indeed the case. One informative way to see this is through the excess free energy per Kuhn monomer due to excluded volume interactions shown in the inset of Fig. 4.4 as a function of contour length. Observe that the excluded volume interactions begin to "turn on" very early, near $L/b \approx 1$, which explains why dsDNA never truly approaches Gaussian scaling. The free energy curve then consumes another three decades in L/b before it nears the asymptotic limit, which further accounts for the broad transition to Flory scaling.

Another consequence of the gradual ramp-up of excluded volume interactions manifests itself through the form factor, which is particularly useful for studying the scaling

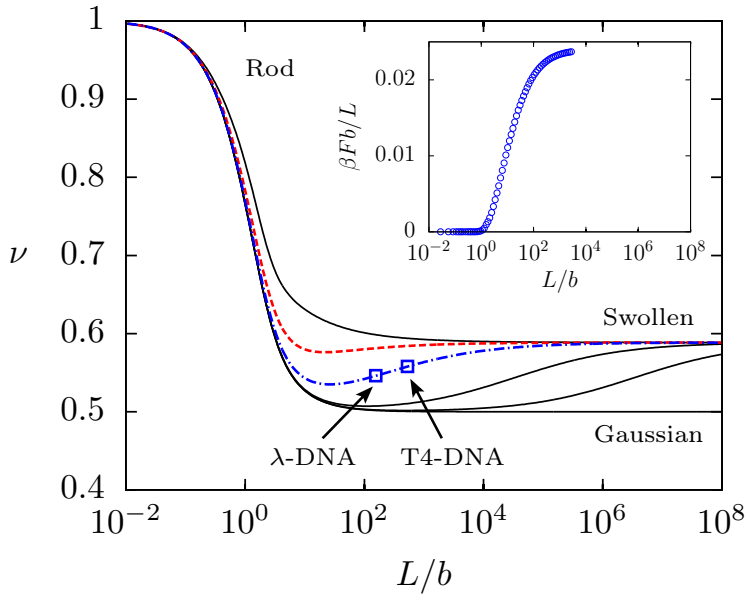

Figure 4.4: Power-law exponent of the end-to-end distance of a semiflexible chain with excluded volume as calculated by results from renormalization group theory [212]. As outlined in Sec. 4.3.1, $\nu = 1$ corresponds to rod-like behavior, $\nu = 0.5$ to a Gaussian chain and $\nu = 0.588$ to a swollen chain. Results are shown for five different values of w/b (from top to bottom): 1.0, 0.217 (ssDNA), 0.043 (dsDNA), 3.16×10^{-3}, 3.16×10^{-4} and 0 (no excluded volume). (Inset) PERM results for the excess free energy per Kuhn length due to excluded volume interactions in a dilute solution of dsDNA.

behavior. The form factor is not only directly available from light scattering experiments [72], but as the Fourier-transform of the pair-correlation function, it provides information about a variety of length scales of a given polymer as a function of the wave vector. In particular, we are interested in the so-called fractal regime $qS \gg 1$ where q is the magnitude of the wave vector and S is the radius of gyration. This regime provides information related to the chain statistics inside the coil, which can reveal details about stiffness and self-avoiding behavior.

The expression for the form factor of a Gaussian chain is the well-known Debye equation

$$P(q) = \frac{2}{(qS)^4}\left[\exp\left(-q^2S^2\right) - 1 + q^2S^2\right] \tag{4.21}$$

Since a wormlike chain includes small length scale effects that are unaccounted for by

the Gaussian chain model, the Debye expression is not always valid in the fractal regime. There is a (somewhat complicated) analytical expression for the form factor of an ideal wormlike chain (see online supporting information as well as Spakowitz and Wang [224]). The basic result mirrors scaling theory; length scales where $ql_p \ll 1$ behave like coils and agree with Eq. 4.21, whereas length scales where $ql_p \gg 1$ behave rod-like and disagree with Eq. 4.21.

The problem becomes more complicated when we include the effect of excluded volume. Although, a closed-form expression for the form factor of a wormlike chain with excluded volume is not available, it can be computed numerically [172, 187, 225]. Additionally, Sharp and Bloomfield [226] have provided a semi-empirical relation (for additional details see online supporting information). However, scaling again helps us interpret the anticipated results. When $z \ll 1$ coils should scale like $P(q) \sim q^{-2}$, which agrees with the Debye equation, and when $z \gg 1$ the form factor should scale like q^{-1/ν_F}.

Fig. 4.5 shows the form factor of several different lengths of dsDNA in the fractal regime using PERM. In this region in particular, long chains such as λ-DNA show a deviation from Eq. 4.21 due to excluded volume effects. This can be seen by the gradual transition from a slope of -2 for short chains (which agrees well with the Debye expression) to a slope of $-1/\nu_F$ for long chains.

The excluded volume dependence of the form factor is of particular importance when extracting the radius of gyration or persistence length from light scattering measurements [95, 181, 213, 226, 227]. This is supported by Fig. 4.5, which shows that a systematic bias in the fitting parameters (*i.e.* the radius of gyration) is present if Eq. 4.21 is used to fit the PERM data for long chains. We speculate this principle provides an explanation for the contradiction between the radius of gyration extracted from classic light scattering studies on T7 DNA (40 kbp) [181], which showed excluded volume effects, and recent fluorescent correlation spectroscopy measurements of chains up to 97 kbp [180], which did not. Accordingly, we conclude that one must resort to either simulations or the semi-empirical relation of Sharp and Bloomfield to accurately estimate the size of long dsDNA.

Thus far we have discussed two principles that emerge from a quantitative evaluation of the equilibrium properties of DNA (and semiflexible chains in general). Namely,

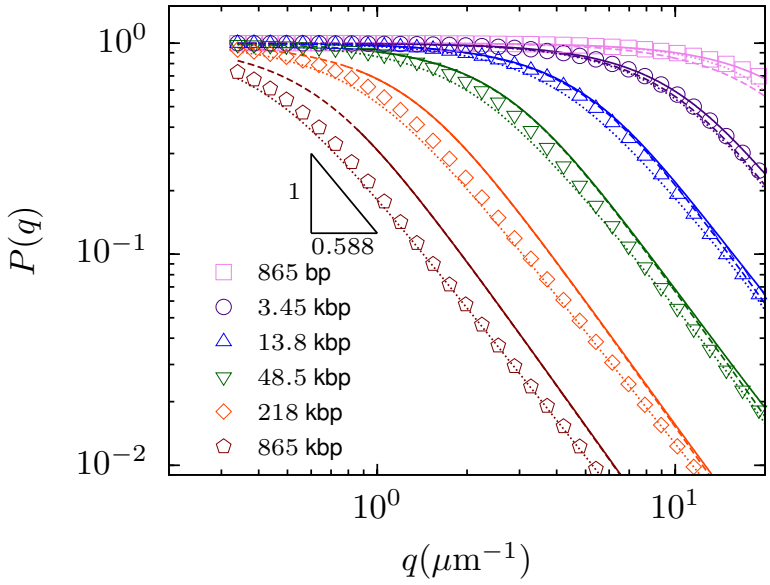

Figure 4.5: Form factor of dsDNA for $ql_p < 1$ for various contour lengths (from left to right): 865 kbp, 218 kbp, 48.5 kbp (λ), 13.8 kbp, 3.45 kbp, 865 bp. Symbols correspond to PERM calculations (with excluded volume), dashed lines to Eq. 4.21 and dotted lines to a semi-empirical expression by Sharp and Bloomfield [226]. Solid lines correspond to the form factor of an ideal wormlike chain with the same stiffness [224] (see online supporting information).

transitions are smooth and asymptotic and excluded volume effects are important at length scales well below the flexible limit. Additionally, simulation results show that different size metrics have quantitatively different transitions. Since scaling theory is unable to predict such transitions, this has been underappreciated in the literature. In fact, when combined with smooth and broad transitions, metric-dependent transitions make it very difficult (if not impossible) to define an objective measure of when a wormlike chain like DNA is "in a regime."

To see this, consider Fig. 4.6 which depicts the scaling exponent ν for the size metrics S and R from both PERM and RG calculations as well as X, which is available from PERM alone. Due to high-frequency fluctuations in the PERM data, the derivative ν was determined by a Savitsky-Golay filter with second-order polynomials. Even with the filter, some low-frequency noise still exists in the data, causing fluctuations at large

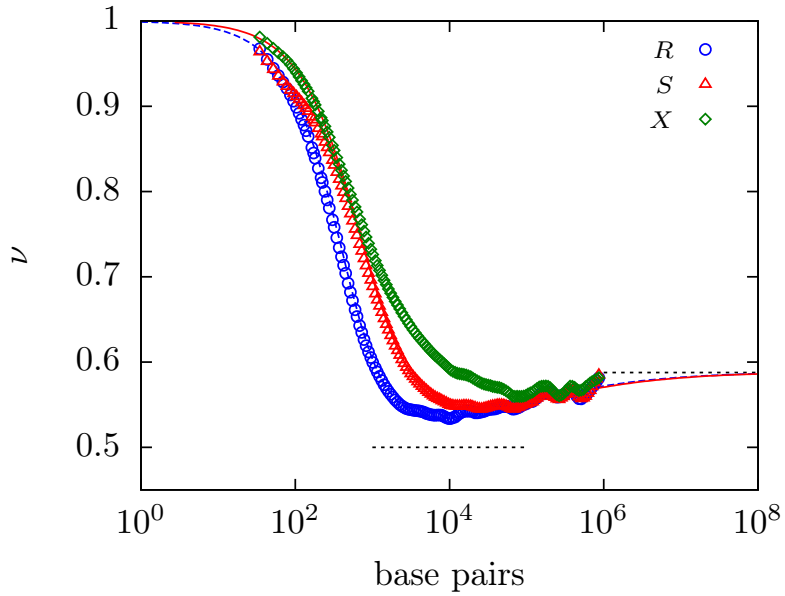

Figure 4.6: PERM calculations of dsDNA for the value of the exponent ν in $R \sim L^{\nu}$ (circles), $S \sim L^{\nu}$ (triangles) and $X \sim L^{\nu}$ (diamonds) for dsDNA ($w/b = 0.0434$) and RG calculations for ν for R (dashed line) and for S (solid line). The dashed horizontal lines correspond to the values of $\nu = 0.5$ (Gaussian) and $\nu = 0.5876$ (swollen).

molecular weights. Nevertheless, the PERM results for the radius of gyration and end-to-end distance agree very well with the RG theory calculations until very small L, which is caused by the discretization of the DWLC model.

Fig. 4.6 clearly shows that each metric has a different minimum and a different approach to the long-chain limit. For instance, it appears that the end-to-end distance has the deepest minimum and the longest climb to the asymptotic limit whereas the mean span has a very shallow minimum. Consequently, with finite contour lengths, any computation or measurement is going to exhibit metric dependent behavior. In other words, measurements with the same molecular weight of dsDNA with different metrics will result in different scaling exponents. Furthermore, it appears for some molecular weights, that the measured scaling exponent ν may be more sensitive to the size metric than to the contour length.

The dependence of the scaling exponent on the size metric also illuminates the concept of the thermal blob. Briefly introduced in Sec. 4.3.1, the thermal blob can be

understood as a renormalized monomer in a flexible, self-avoiding chain. In other words, a very long, flexible chain can be viewed as a self-avoiding walk of thermal blobs [72]

$$\frac{\Xi}{\xi_{\mathrm{T}}} \sim \left(\frac{L}{l_{\mathrm{T}}}\right)^{\nu_{\mathrm{F}}} \tag{4.22}$$

where Ξ is some size metric (e.g. radius of gyration), L is the total contour length and ξ_{T} is the size of a blob composed of a subsection, l_{T}, of the contour length of the original chain. Therefore to be truly flexible and self-avoiding, the contour length of a chain must be much greater than the blob length and its size must be much greater than the blob size.

Since the thermal blob length is a scaling parameter, it is useful as a qualitative measure only, and consequently, one should not expect a single value of l_{T} to provide a precise value of the minimum length-scale for excluded volume interactions. This enters explicitly in the definition of the thermal blob length l_{T} in Eq. 4.18 where the constant c was left undefined. Nevertheless, there are several definitions of c which are commonly encountered in the literature and we find it useful in Table 4.2 to make a comparison of the contour length and size of the resulting thermal blob obtained using these definitions.

As a first estimate, one may simply set $c = 1$, which gives $l_{\mathrm{T}} = 64$ bases for ssDNA and $l_{\mathrm{T}} = 166$ kbp for dsDNA. One may also set the excluded volume parameter z equal to one, which leads to $c = (2\pi/3)^3$ and consequently $l_{\mathrm{T}} = 587$ bases for ssDNA and $l_{\mathrm{T}} = 1.52$ Mbp for dsDNA. A more rigorous definition sets the thermal blob length to be the contour length of a chain where the excess free energy from excluded volume is equal to $k_B T$ [72, 228]. Since PERM directly computes the excess free energy, this length is immediately available and yields $c = 0.102$ for dsDNA (16.8 kbp) and $c = 0.458$ for ssDNA (45 bases).

Even for a fixed value of c, the thermal blob size, ξ_{T}, also depends on the size metric. Table 4.2 shows one such example using $F = k_B T$ to pick c. Here, with a fixed contour length of $l_{\mathrm{T}} = 16.8$ kbp, the radius of gyration of a thermal blob of dsDNA is 332 nm, but the end-to-end distance is 825 nm. Similarly, for $l_{\mathrm{T}} = 45$ bases, the radius of gyration of a thermal blob of ssDNA is 3.9 nm, but the end-to-end distance is 10.2 nm.

While it is not surprising that l_{T} and ξ_{T} vary with the choice of c and size metric, the magnitude of the variation that is represented in Table 4.2 is somewhat startling.

Table 4.2: Thermal blob length l_{T} and size ξ_{T} of ssDNA and dsDNA determined using different choices of scaling constant c and different size metrics.

	l_{T}	
	ssDNA	dsDNA
$c = 1$	64 bases	166 kbp
$z = 1$	587 bases	1520 kbp
$F = k_B T^*$	45 bases	16.8 kbp
	ξ_{T} $(F = k_B T)$	
	ssDNA	dsDNA
End-to-End Distance	10.2 nm	825 nm
Radius of Gyration	3.9 nm	332 nm
Mean Span	7.3 nm	678 nm

*The thermal blob length and size for the $F = k_B T$ case were computed using PERM calculations for the excess free energy and the various size metrics respectively.

Given reasonable but different choices of the constant c in Eq. 4.18, we find that the thermal blob length can vary by nearly two orders of magnitude and encompasses much of the range of molecular weights available for experiments. The large variation in the thermal blob lengths in Table 4.2 further emphasizes their qualitative nature and cautions that these values should only be considered rough order of magnitude estimates of the length-scale where excluded volume and bending effects are approximately equal. Finally, this variation suggests that a direct computation of "how many thermal blobs are in a chain" is not meaningful without a specific definition of c. For instance, using the various definitions of c found in Table 4.2, λ-DNA can be reported to have 2.9 ($F = k_BT$), 0.29 ($c = 1$), or 0.032 ($z = 1$) thermal blobs. In contrast to these estimates, when prefactors and transitions are fully resolved, we unambiguously show in Fig. 4.4 and Fig. 4.6 that λ-DNA is in the middle of a transition from ideal to Flory scaling.

4.3.4 Dynamic Properties of DNA

To this point, the discussion has focused on the equilibrium properties of DNA and the role of excluded volume as a wormlike chain approaches the flexible chain limit. However, the near-equilibrium diffusion coefficient and other dynamic properties also play a prominent role in the use of dsDNA as a model polymer. The effects of solvent mediated interactions between distal chain segments, called hydrodynamic interactions (HI), are central to dilute solution DNA dynamics. Hydrodynamic interactions are in many ways the dynamic counterpart to excluded volume interactions, and their inclusion introduces a degree of freedom through the hydrodynamic diameter d.

Despite some similarities, the hydrodynamic diameter is not in general equal to the effective width w, since chain friction and excluded volume arise from distinct physical phenomena. In principle, the hydrodynamic diameter corresponds to the surface of shear of the molecule and is an intrinsic property of a polymer chain. However, in our case we have employed a far-field approximation (Eq. 4.13) that neglects near-field lubrication forces, rendering the hydrodynamic diameter a phenomenological parameter.

With the addition of a degree of freedom for the hydrodynamic interactions, the landscape of possible types of diffusive behavior for a wormlike chain becomes complicated [128]. The diffusion coefficient depends not only on configurational properties (including excluded volume), but also on the strength of the HI. In other words, there

is not a simple one-to-one correspondence between configuration and diffusive behavior (even for very flexible chains). The literature identifies at least three classes of behavior that a wormlike chain can exhibit, which are shown in Fig. 4.7.

For very short and stiff conformations with HI, the chain exhibits rod-like diffusion [114, 129, 229]

$$D_{\text{rod}} = \frac{k_B T}{3\pi\eta L}\left[\ln(L/d) + \gamma\right] \tag{4.23}$$

where η is the solvent viscosity and γ is in general a function of L/d, and equals $2\ln 2 - 1$ in the slender-body limit. For more flexible chains, one can distinguish between the case where HI is weak (free-draining) and the case where HI is strong (non-draining) [126]. As indicated in Fig. 4.7, a free-draining coil experiences no hydrodynamic screening and interacts fully with the solvent giving the Rouse diffusion coefficient [114]

$$D_{\text{Rouse}} = \frac{k_B T}{6\pi\eta L} \tag{4.24}$$

The non-draining coil has significant HI and is impermeable to solvent flow. For a flexible chain with no excluded volume interactions, the diffusion coefficient was derived by Zimm [114, 126]

$$D_{\text{Zimm}} = \frac{8}{3\sqrt{6\pi^3}}\frac{k_B T}{\eta\sqrt{Lb}} \tag{4.25}$$

and gives a diffusion coefficient that is inversely proportional to the coil size.

While it is clear that a wormlike polymer may exhibit any one of these classes of behavior, a comprehensive qualitative description of the problem for wormlike chains remains elusive. Indeed the inclusion of the effects of flexibility, excluded volume and hydrodynamic interactions has proven to be an exceedingly difficult task [90, 127–129, 183] and accordingly, no complete analytical theory exists to date. However, several important pieces have been developed since the work on flexible polymers by Zimm [126], some of which are noteworthy.

First, Oono and Kohmoto [114, 127] used a dynamic renormalization group theory to find the diffusivity of a flexible polymer chain with both EV and HI. In agreement with the result by Zimm, the diffusivity

$$D_{\text{Oono}} = \frac{1}{12.067}\frac{k_B T}{\eta S} \tag{4.26}$$

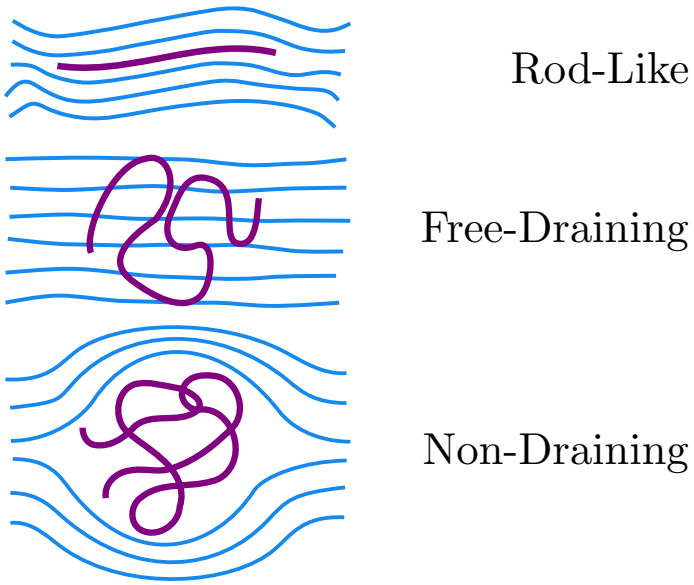

Figure 4.7: Schematic of diffusive behaviors of wormlike chains. Rod-like diffusion dominates for very short, stiff chains. Polymer coils can exhibit either free-draining (weak HI) or non-draining (strong HI) behavior. Partial draining behavior is also possible for chains with relatively open structures. In this case, the polymer conformation is not sufficient to describe the diffusive behavior of the chain since the strength of the HI (*i.e.* the hydrodynamic radius) can vary independently.

suggests that the long-chain limit is characterized by non-draining coils where the dynamics are governed by conformational effects only (the hydrodynamic diameter is conspicuously absent).

However, further work by Douglas and Freed [128] displays a complicated picture for finite-length chains where excluded volume effects (which swell the coil) act in competition with hydrodynamic interactions which decay with decreasing chain density. This competition can lead to *partial-draining*, an intermediate state between the free-draining and non-draining limits pictured in Fig. 4.7, where the configurational properties of the chain are insufficient to completely describe the diffusivity. Thus even very long (but finite) chains may not obey Eq. 4.26, but will include a dependence on the hydrodynamic diameter.

In an orthogonal attempt, Yamakawa and Fujii [129] computed the diffusivity of an ideal wormlike chain, which accounts for chain stiffness and HI, but neglected the effects of excluded volume (see supporting information online). The wormlike chain diffusivity shows a gradual crossover from rod-like behavior (Eq. 4.23) to Zimm diffusion (Eq. 4.25) as the contour length increases, which in turn means a gradual decrease in the effect of the hydrodynamic diameter. Given that our DWLC model incorporates all of the effects listed above, we anticipate that the diffusion of DNA will include effects from each of these previous works.

To correctly capture the dynamics, we need an accurate estimate for the hydrodynamic radius d of ssDNA and dsDNA. Literature values for the hydrodynamic diameter of dsDNA have been obtained by a variety of experimental methods and typical values vary between 2 to 3 nm [31, 83–85, 87–89, 92–94, 96, 97, 99, 101, 104–108, 129, 230, 231]. Less is known about the hydrodynamic radius of ssDNA. The diffusivity of short ssDNA chains can be sequence dependent due to base-pair interactions thus requiring thermal and chemical denaturing agents [222], which makes generic diffusivity studies difficult [179]. Due to a lack of data therefore, the hydrodynamic radius is uncertain, and we simply assume that $d = w$.

To get a more precise value of d for dsDNA, Fig. 4.8 shows a meta-analysis of several measurements of the diffusivity of dsDNA in the literature [83–85, 87–89, 92–94, 96, 97, 99, 101, 104–108]. Here, the Kirkwood diffusivity for dsDNA is plotted alongside the experimental data as a function of molecular weight, which is rescaled by the Zimm

diffusion given in Eq. 4.25. Note that Eq. 4.25 is not defined in terms of the radius of gyration, S, of the polymer, but rather in terms of the contour length, L, and Kuhn length, b. This allows an unambiguous comparison to a larger experimental data set since there is only a small overlap between the sources of experimental data for the diffusivity (Fig. 4.8) and the radius of gyration (Fig. 4.3). The hydrodynamic radius is extracted by comparing the experimental data to the theory of Yamakawa and Fujii [129] for the diffusion of a wormlike chain without excluded volume. This is justified since the diffusivity is most sensitive to the hydrodynamic radius at low L where hydrodynamic screening and excluded volume interactions are negligible. As seen in Fig. 4.8, the value of 2.9 nm (which agrees with an analysis by Lu et al. [231]) fits the low molecular weight data exceptionally well.

Fig. 4.8(B) also shows that at large contour length, the PERM diffusivity calculations give excellent agreement with values of the diffusion obtained from dynamic light scattering (DLS) and sedimentation experiments. The agreement with the experimental data at low molecular weight is expected, since it was used to obtain the hydrodynamic radius. However, the agreement with the DLS and sedimentation data persists for large molecular weights, when excluded volume effects cause the chain to swell. This suggests that the both the size and degree of hydrodynamic screening of the dsDNA coil is well described by the DWLC model.

In contrast, the DWLC model does not agree well with single molecule diffusivity measurements from fluorescence microscopy [89, 92], which we hypothesize is due to the presence of intercalating dyes. It is unclear how the width, persistence length and hydrodynamic radius of DNA change with intercalating dyes [185, 186], making it difficult to computationally replicate their effect on diffusivity. Since the contour length of λ-DNA is observed to increase from 16.3 μm to about 21 μm, a common supposition is that all properties increase by a constant factor, leaving the ratios between properties constant. This proposition is easily tested by our model, and assuming an increase of 28%, we find that the change in diffusivity is insufficient to explain the disagreement (see supporting information online). Instead, we conclude that in addition to changing the contour length, fluorescent dyes are likely to alter the ratios b/w and d/w. Indeed, this appears reasonable since the positively charged dyes lead to a decrease in the effective charge of the DNA [232], which may decrease the effective width.

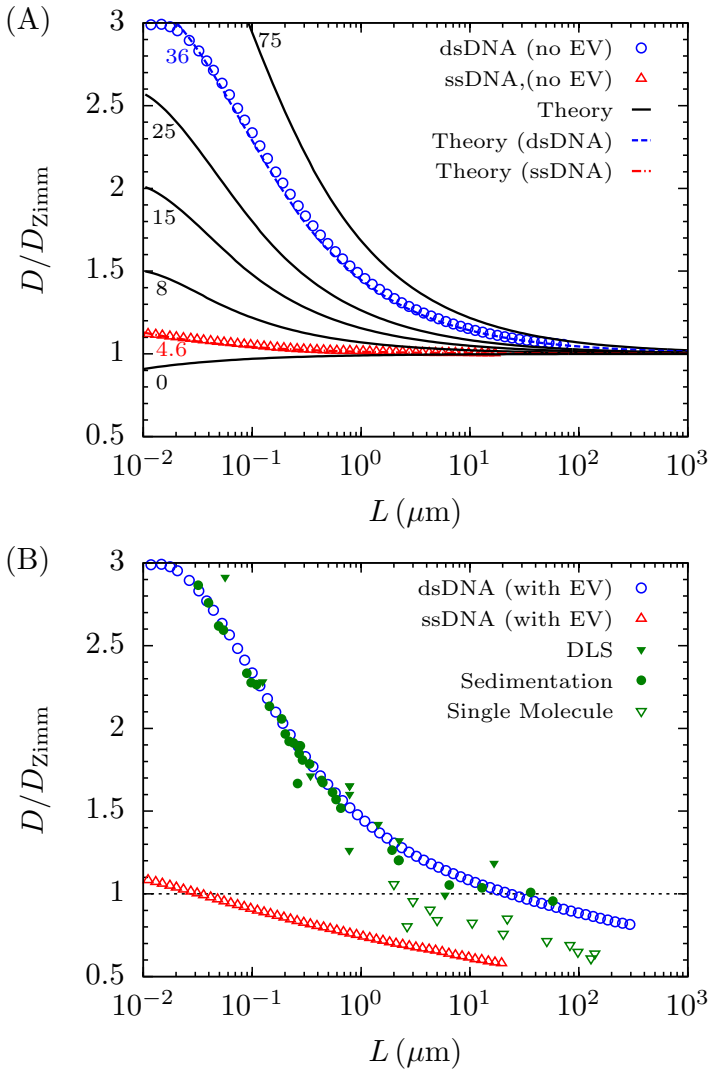

Figure 4.8: Parameterization of the hydrodynamic diameter for the DWLC. (A) Curves (solid, dashed, dash-dot) show the relative diffusivity of a CWLC without excluded volume [129] at different values of b/d. The touching bead DWLC model (symbols) shows excellent agreement with the CWLC chain (curves). (B) Experimental data for the diffusivity from dynamic light scattering (filled triangles) [83, 93, 94, 96, 99, 101, 104, 108], sedimentation (filled circles) [84, 85, 87, 88, 97, 105–107], and single molecule methods (open triangles) [89, 92]. The dsDNA data fits the DWLC simulation (now with excluded volume interactions) for $b/d = 36$ or $d = 2.9$ nm. Notice that since the diffusivity is scaled by the Zimm diffusion (ideal chain diffusion), the asymptotic value does not approach 1. Additionally, it appears that the single molecule data give poorer agreement with the simulation data, presumably due to the impact of intercalating dyes.

Moving forward, we would like to examine the behavior of ssDNA and dsDNA as the diffusivity approaches the long chain limit. Notice that if we introduce the definition of the chain hydrodynamic radius

$$R_{\mathrm{H}} \equiv \frac{k_B T}{6\pi\eta D} \tag{4.27}$$

The ratio of the radius of gyration to the hydrodynamic radius becomes [127]

$$\frac{S}{R_{\mathrm{H}}} = 1.562 \frac{D}{D_{\mathrm{Oono}}} \tag{4.28}$$

In the long-chain limit, $D \to D_{\mathrm{Oono}}$ and the ratio is predicted to converge to a universal value of 1.562 [127].

Fig. 4.9(A) shows the ratio S/R_{H} as a function of molecular weight. As expected, both curves appear to approach a constant value in the limit that $L \to \infty$, but the value appears slightly larger than predicted by the renormalization group theory. (A least squares fit to the ssDNA data for $L > 2000$ bases gives $S/R_{\mathrm{H}} = 1.58902(2)$.) For ssDNA, the story is much the same as it was for static properties; within 100 bases, the diffusion appears to have reached its long-chain limit. Also similar to the static size measures, it takes an exceptionally long dsDNA chain to reach the flexible coil limit. According to Fig. 4.9(A), dsDNA is within about 1% of the value predicted by Oono and Kohmoto [127] by the terminal molecular weight of 865 kbp.

The claim that dsDNA converges very slowly to its long-chain value is at odds with earlier experimental work which asserted that λ-DNA is fully swollen and non-draining [89, 92]. However, the evidence for non-draining coils is based upon the measured scaling exponent ($\nu = 0.61$), which appears to be relatively unresponsive to the draining behavior of dsDNA (see supporting information online). Experimental work by Schroeder et al. [182, 233] made more sensitive measures of the hydrodynamic behavior of dsDNA by studying the phenomena of conformation hysteresis. Schroeder et al. found that significant hydrodynamic effects were observed for chains near 1.3 Mbp, and that it took a nearly 3 Mbp polymer to achieve the sought-after hysteresis. While these experiments did not search for the onset of HI interactions per se, the effects of HI are observed in chains with molecular weights within an order of magnitude of those predicted by our calculations. Subsequently, their observations support our conclusion that extremely long chains are needed to achieve non-draining behavior.

The slow convergence to non-draining behavior also has practical implications for the measurement of the radius of gyration in fluorescence microscopy experiments. It is common practice to use fluorescence microscopy measurements of the diffusivity to estimate the radius of gyration by assuming that the $D/D_{\mathrm{Oono}} = 1$ in Eq. 4.28 [89, 92]. As shown by Fig. 4.9(A) there is always some systematic bias made by this inference. However, the bias decreases as molecular weight increases making the assumption justified at very large molecular weights. Assuming that the constant 1.562 is exact (which is questionable), this method underestimates the radius of gyration of λ-DNA by about 9%.

The extremely large molecular weights required to reach the non-draining limit also bring to mind the previous discussion surrounding partial draining. To further understand the partial draining of dsDNA, consider Fig. 4.9(B), which shows the normalized Kirkwood diffusivity as a function of molecular weight. Here the diffusivity is shown to cross over from rod-like diffusion (Eq. 4.23) at low molecular weights to non-draining diffusion (Eq. 4.26) at large molecular weights. According to Fig. 4.9(B), dsDNA less than a few hundred base pairs is well approximated by rod-like diffusion (to within a constant factor). This seems reasonable, given that a chain of 156 bp is about one persistence length. As the contour length increases, the diffusivity is observed to asymptotically approach the non-draining limit. Consistent with the previous analysis, this asymptotic approach is slow, and λ and T4 DNA give diffusivities that are respectively 9% and 4% greater than the asymptotic limit. We conclude therefore, that kilobase-pair length dsDNA is partially draining.

While partial draining has been a subject of discussion since at least 1979 (couched in terms of dynamic scaling [77]), it has become a topic of recent interest in both free solution [117, 183] and confinement [154, 155] (see also Chapter 6). Interestingly, Mansfield and Douglas [183] suggest that transport properties are especially slow to converge to their long-chain values. Their work, which employs several different polymer models, suggests that a slow transition to non-draining behavior is not unique to DNA. This trend is also seen in recent work by Dai et al. [155] on the diffusion of DNA in slits, which posits that the local pair correlation function of a polymer has a long-ranged impact on coil dynamics. Unfortunately, due to noise in the present data, it is difficult to tell whether or not the diffusion coefficient converges more slowly than the radius of

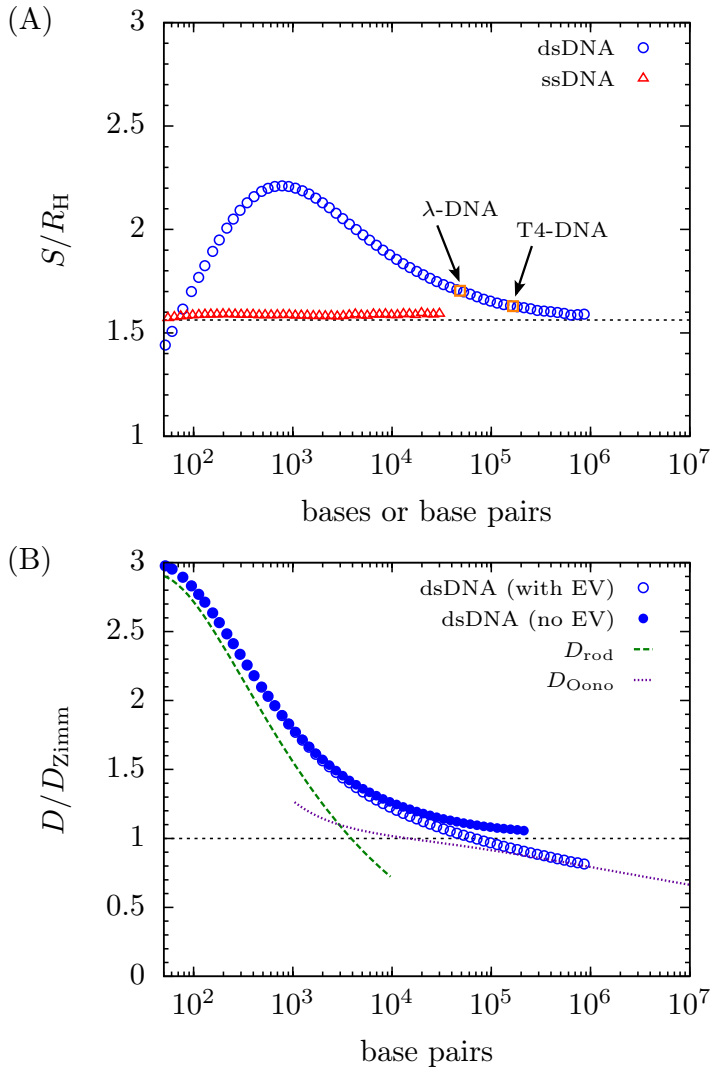

Figure 4.9: (A) PERM results for the ratio of the radius of gyration S to the hydrodynamic radius R_H as a function of the molecular weight for both ssDNA (open triangles) and dsDNA (open circles). λ-DNA and T4-DNA are shown for reference (arrow, open squares). The horizontal dashed line corresponds to $S/R_H = 1.562$. (B) The diffusion coefficient from PERM, rescaled by Eq. 4.25, as a function of the number of base pairs of DNA. Double-stranded DNA is shown to transition from rod-like behavior (dashed line) to non-draining behavior (dotted line) over several orders of magnitude in molecular weight. The diffusion coefficient without excluded volume (closed circles) is shown for reference.

gyration or other static measures.

4.4 Conclusion

Using a powerful Monte Carlo method, PERM, we have elucidated the long-chain behavior of both single-stranded and double-stranded DNA. Clearly, single-stranded DNA is much more flexible and several hundred bases are sufficient to guarantee both complete swelling and non-draining behavior. By contrast, double-stranded DNA is much slower to reach flexible-chain behavior. It appears that dsDNA much less than 1 Mbp should not be considered either completely swollen or fully non-draining. One immediate consequence of this result concerns the practice of inferring the radius of gyration of dsDNA from the diffusivity, which we find leads to a systematic bias (underestimation) in the measurement of the radius of gyration.

In addition, we find that shorter chains (*e.g.* λ-DNA), while not completely swollen, are nevertheless influenced by excluded volume and hydrodynamic interactions. This is complicated by the fact that the transitions between universal regimes are continuous and the approach is asymptotic and metric specific. Combined together, these observations suggest that it is inappropriate to consider λ-DNA as an "ideal" chain and neglect EV and HI. In some sense, λ-DNA is possibly the worst model polymer since it is clearly in a transition.

On the upside, even though the excluded volume and hydrodynamic interactions are significant, the languid transition towards universal behavior also indicates that the measured properties of dsDNA do not change rapidly as a function of contour length. In other words, practical estimates of the radius of gyration and diffusion coefficient are relatively unaffected by the change in scaling exponent as long as the change in molecular weight is not too great over the range of the estimate. Accordingly, Brownian dynamics studies that do not account for HI or EV explicitly, can in principle reproduce properties quantitatively by careful parameterization. However, such a parameterization will only be valid for a small range of contour lengths, and the properties obtained from subsequent simulations should be limited to this range. Nevertheless, prudence is warranted in interpreting such simulations, since the correct physics is not inherently incorporated.

Certainly, one troubling implication of our results concerns the lack of agreement with data from fluorescence microscopy experiments. We have attempted to justify this by the presence of fluorescent dyes, which alter the backbone of dsDNA. We believe that more work is needed to account for the disagreement and propose that the effects of intercalating dyes on dsDNA be studied in greater detail. In addition to this, we have not considered here the effect of changing ionic strength, which should be straightforward with minor modifications to the DWLC model.

In conclusion, we find it difficult to give a straightforward answer to the question: Is DNA a good model polymer? On the one hand, dsDNA continues to be widely used due to its extraordinarily useful experimental properties. Among others these include exquisite contour length selectivity, near monodisperse solutions, direct visualization techniques, ideal size relative to nanofabricated devices and end-attachment chemistry. On the other hand, if we narrow our scope to strictly universal behavior, megabase length chains appear to be necessary. Such long contour lengths give rise to experimental difficulties including chain cleavage and knotting, and at the present moment, it appears that such large chains are atypical in polymer dynamics experiments. Given this, we recommend caution in the interpretation of dynamic data obtained by shorter dsDNA.

Chapter 5

Regimes of DNA confined in Nanochannels

This chapter is based on the publication:
D. R. Tree, Y. Wang, and K. D. Dorfman
Phys. Rev. Lett. 110 (2013), 208103

5.1 Introduction

When a wormlike polymer such as DNA is confined in a long channel whose width is smaller than the polymer's free solution radius of gyration, steric interactions with the walls cause the polymer to extend along the channel axis. The classical theories describing this phenomenon, sketched in Fig. 5.1, were described by Odijk [58] and de Gennes [59], respectively, over 30 years ago. However, these theories are only valid in the impractical cases of very strong ($D \ll l_p$) or very weak ($D \gg l_p$) confinement, respectively, where D is the channel size and l_p is the persistence length of the chain. In this Letter, we establish that the relevant intermediate regime for DNA extension in a nanochannel is a universal de Gennes-like regime with ideal blobs. We arrived at this conclusion by recognizing the connection with the rod-to-coil transition for free wormlike polymers [73, 172, 209, 212, 234] illustrated in Fig. 5.1. In addition to describing the experimentally relevant phenomena for DNA, this connection leads to a complete description of the universal regimes for all long channel-confined wormlike chains that we validated

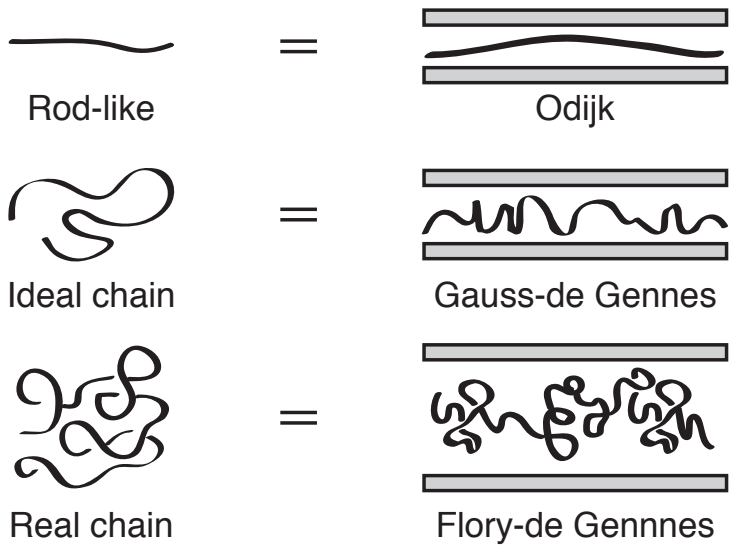

Figure 5.1: Illustration of the analogy between free solution and confined configurations of a wormlike chain. The classical theories renormalize the chain into a series of subchains, where these subchains are either rod-like (Odijk) or excluded volume blobs (de Gennes). We demonstrate here the existence of a universal Gauss-de Gennes regime in confinement that connects the (rod-like) Odijk and (excluded volume) de Gennes regime. For clarity, we refer to the classic de Gennes regime as the "Flory-de Gennes" regime to highlight its excluded volume nature.

using large-scale Pruned-Enriched Rosenbluth Method (PERM) simulations [171, 173].

To understand the analogy in Fig. 5.1, let us first recall the regimes of chain conformations in free solution. Three regimes characterize the normalized end-to-end distance of a wormlike chain, $\rho \equiv \langle R^2 \rangle^{1/2}/l_p$. This distance depends on two dimensionless numbers: $N \equiv L/l_p$, the number of persistence lengths in a chain of length L, and an anisotropy parameter $\epsilon \equiv w/l_p$, which measures the relative strength of the excluded volume interactions (quantified by the effective chain width w) to the bending energy. In the limit of negligible excluded volume interactions [72], $z \equiv \epsilon N^{1/2} \ll 1$, the Benoit-Doty equation for a continuous wormlike chain gives [73]

$$\rho^2/2 = N - 1 + \exp(-N) \tag{5.1}$$

This model predicts a stiff chain with $\rho \sim N$ for $N \lesssim 1$ and Gaussian statistics with $\rho \sim N^{1/2}$ for $1 \lesssim N \lesssim \epsilon^{-2}$. For a sufficiently long chain $N \gtrsim \epsilon^{-2}$, excluded volume

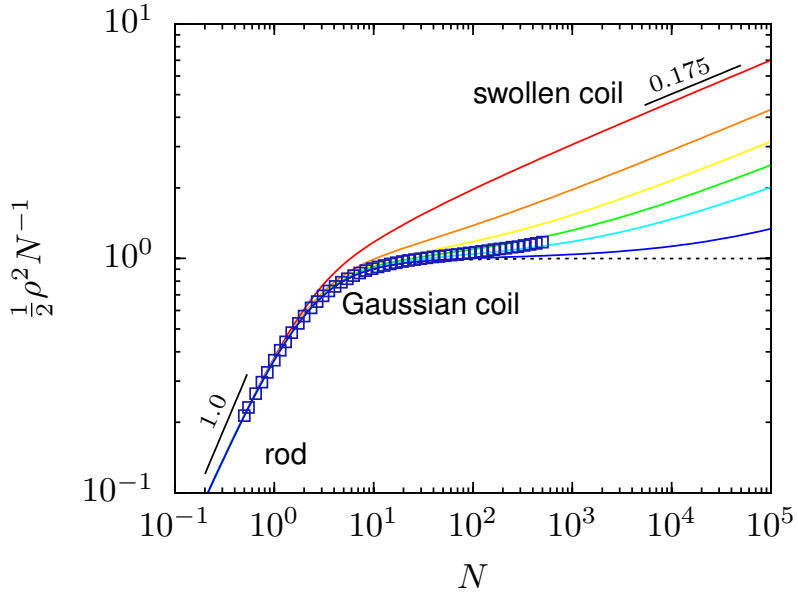

Figure 5.2: The normalized mean-square end-to-end distance of a wormlike chain in free solution as a function of normalized chain length from Eq. 5.1 ($\epsilon = 0$, dashed line), renormalization group theory [212] with (descending from top to bottom in the figure) $\epsilon = 2$, 0.5, 0.2, 0.1 (DNA), 0.05, 0.01, and PERM simulations for $\epsilon = 0.05$ (squares, $L_{\max}/w = 2 \times 10^4$). The most flexible chain corresponds to an upper bound $\epsilon = 2$, where the Kuhn length of the chain equals its width [210].

interactions are important and $\rho \sim \epsilon^{2\nu-1} N^{\nu}$, where $\nu = 0.5877$ is the modern value of the Flory parameter [80]. As evident in Fig. 5.2, the scaling in the Gaussian regime is not exactly that for an ideal chain due to finite excluded volume effects. Moreover, weakly anisotropic chains — such as DNA, which is only a moderately stiff biopolymer ($\epsilon \approx 0.1$) [71] — have a very narrow Gaussian regime. However, in the limit $\epsilon \to 0$, the Gaussian regime spans an infinite amount of chain length and is thus a universal regime. Accordingly, many biopolymers [235] are stiff enough to exhibit broad Gaussian regimes.

We will demonstrate, via simulations, that three similar regimes characterize the confinement free energy of an asymptotically long wormlike chain confined in channel. Following Odijk [58] and de Gennes [59, 77], the chain is renormalized into N/g units containing g persistence lengths per unit. This in turn implies that the chain properties

(confinement free energy and extension) are extensive, as has been shown many times [58, 59, 77, 236] for an infinite chain in a quasi-1D geometry.

5.2 Theory

The Odijk regime [58] in Fig. 5.1 corresponds to rod-like behavior over the length scale D. For channel sizes $\delta \equiv D/l_p \lesssim 1$ [58], the stiff chain projects a distance of $\lambda \sim (D^2 l_p)^{1/3}$ before deflecting off of the walls. This makes the number of persistence lengths in the correlation volume

$$g = \lambda/l_p \sim \delta^{2/3} \tag{5.2}$$

Assuming an energy of $k_B T$ per independent segment [58] gives the dimensionless confinement free energy,

$$\mathcal{F} \equiv \frac{\Delta F_c}{N k_B T} \sim 1/g \sim \delta^{-2/3} \tag{5.3}$$

The extension is given by the projection of the deflection segment length onto the channel axis $X = (N/g)\lambda \cos\theta$ [58] which simplifies to

$$\langle X/L \rangle = 1 - \alpha\, \delta^{2/3} \tag{5.4}$$

where the prefactor $\alpha = 0.18274$ for a square nanochannel is given by high resolution simulations [132]. Analogous to the rod-like behavior in free solution, the thermodynamics of the Odijk regime is independent of the width of the chain.

Continuing with the analogy, the de Gennes regime [59] in Fig. 5.1 corresponds to real chain statistics, which leads us to call it the "Flory-de Gennes" regime. Here, as was the case for real chains in free solution, we need to account for the finite chain width. To do so, we use the concept of a "blob" to denote a section of the chain with g persistence lengths that has a correlation length equal to the channel size D. Recalling that the Flory radius for a chain in a good solvent is $R_F/l_p \approx \epsilon^{2\nu-1} N^{\nu}$ [72], the blobs have the size

$$\delta \approx \epsilon^{2\nu-1} g^{\nu} \tag{5.5}$$

With the assumption that the free energy scales as $k_B T$ per blob [59], we have $\mathcal{F} \sim 1/g$, or

$$\mathcal{F} \sim \left(\delta \epsilon^{1-2\nu}\right)^{-1/\nu} \tag{5.6}$$

Following the same reasoning, the extension $\langle X \rangle$ is also extensive in the number of blobs, $\langle X \rangle \cong (N/g)D$. Substituting Eq. (5.5) in the latter gives the scaling

$$\langle X/L \rangle \cong \delta^{1-1/\nu} \epsilon^{2-1/\nu} \tag{5.7}$$

Since the Flory-de Gennes regime corresponds to the onset of excluded volume interactions [138], we would expect this regime to start when the excluded volume parameter for a blob reaches

$$z_{\text{blob}} \equiv \epsilon g^{1/2} \approx 1 \tag{5.8}$$

We thus find that $g \gtrsim \epsilon^{-2}$ corresponds to the Flory-de Gennes regime. Recall that the excluded volume scaling in free solution begins when $N \gtrsim \epsilon^{-2}$. We thus infer that g in confinement is the analogue of N in free solution. Additionally we note that by combining Eqs. (5.5) and (5.8), we can find the boundary of the Flory-de Gennes regime limit in terms of the channel size, $\delta \gtrsim \epsilon^{-1}$, which proves more useful since the channel size is an experimental observable.

For intermediate channel sizes $1 \lesssim \delta \lesssim \epsilon^{-1}$, the g persistence lengths inside D^3 exhibit approximately Gaussian statistics. The derivation of the confinement free energy follows that for the Flory-de Gennes regime with $\nu = 1/2$, leading

$$\mathcal{F} \sim \delta^{-2} \tag{5.9}$$

Since this regime consists of blobs with Gaussian statistics, we refer to it as the "Gauss-de Gennes" regime. This free energy scaling is the same as that of a channel-confined phantom chain originally derived by Cassassa [59, 236]. As is the case in free solution, the scaling of $\mathcal{F}$ for chains with a finite value of ϵ will not be exactly equal to Eq. (5.9). This arises from the weakness (rather than absence) of excluded volume at the persistence length scale.

In the Gauss-de Gennes regime, the intra-polymer correlations are screened at the channel wall [77, 237], which gives $g \sim \delta^2$ persistence lengths per correlation length, D. Since the extended chain consists of N/g such correlation lengths, the corresponding fractional extension is

$$\langle X \rangle / L \sim \delta^{-1} \tag{5.10}$$

The latter scaling has been observed in a number of previous simulations (see [71, 139, 140, 142, 238] and supporting Fig. S6), but its origin and universal nature (or lack

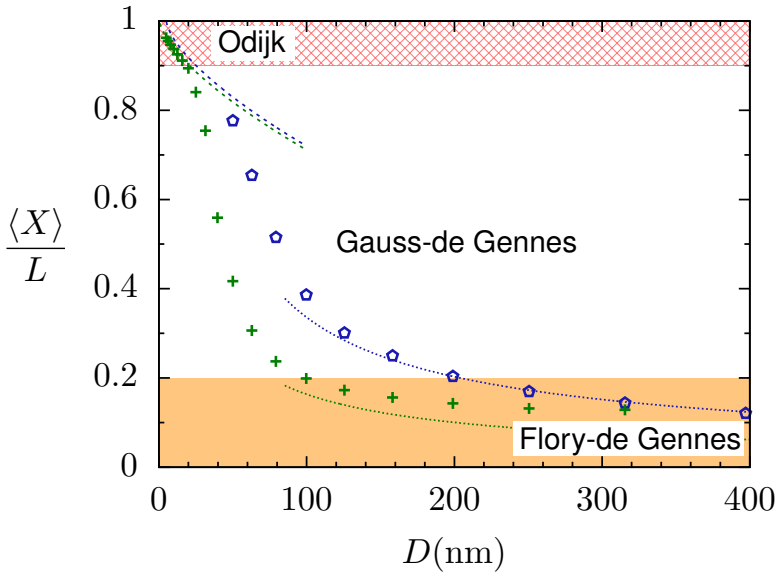

Figure 5.3: Comparison of the fractional extension of the chain predicted by Odijk [58] and de Gennes [59] and simulations of an asymptotically long DNA chain (pentagons, l_p = 50 nm, w = 5 nm, $\epsilon = 0.1$) using the Pruned-Enriched Rosenbluth Method (PERM). The extent of the Gauss-de Gennes regime increases for more filamentous chains (plusses, l_p = 50 nm, w = 0.5 nm, $\epsilon = 10^{-2}$). The shading corresponds to the regimes for $\epsilon = 0.1$.

thereof) have been elusive until now because DNA is not an especially stiff biopolymer. Given this fact one may be tempted to dismiss the regime as unimportant, but consider the case of DNA in a high ionic strength buffer ($\epsilon = 0.1$) which is highlighted in Fig. 5.3. Although the Gauss-de Gennes regime spans less than a decade in dimensionless channel size, these sizes encompass the typical channels used in experiments [30, 31, 47, 55, 60, 237]. Morover, the Flory-de Gennes regime corresponds to $\lesssim 20\%$ extension and the Odijk regime corresponds to $\gtrsim 90\%$ extension, leaving the Gauss-de Gennes regime to span a significant portion of the practically relevant range of fractional extensions for genomic mapping. However, this regime is not in principle limited to a small range of channel sizes. For stiff enough chains, the range of applicable channel sizes $1 \lesssim \delta \lesssim \epsilon^{-1}$ will span many decades, showing the existence of a universal regime.

5.3 Results and Discussion

We have tested this scaling theory in square channels using Pruned-Enriched Rosenbluth Method (PERM) simulations of asymptotically long chains that are long enough to suppress any end effects. PERM is a biased chain-growth Monte Carlo algorithm originally introduced for lattice chains by Grassberger [171]. In the algorithm, "tours" of chains are grown and the Rosenbluth weight of the chain is controlled by selective pruning and enrichment, thus overcoming the attrition problem for the Rosenbluth-Rosenbluth chain growth algorithm [168] for long self-avoiding chains. Choosing efficient parameters for executing the original PERM algorithm is somewhat of an art, and we have followed a parameterless version by Prellberg [173] that simplified the calculation considerably. Our optimized implementation of PERM (see supporting information) allowed us to sample long chain lengths (typically 2×10^4 touching beads of size w) while spanning four decades in the dimensionless channel size δ and three decades in the anisotropy ϵ. For DNA with $w = 5$ nm, our data typically correspond to contour lengths of 100 μm, a full order of magnitude longer than traditional Markov chain Monte Carlo techniques [71, 139, 140, 142, 237, 238]. This combination of asymptotically long chains, a thorough exploration of the (δ,ϵ) phase space (see supporting information), and the large range of confinement free energies allowed us to draw meaningful conclusions about universality. In addition to providing the chain conformational properties shown in Fig. 5.2 and Fig. 5.3, PERM can provide thermodynamic properties like the confinement free energy.

To clearly see the analogy with the rod-to-coil transition in free solution, we also need the equivalent of Eq. (5.1) for the confinement free energy of an ideal wormlike chain ($\epsilon = 0$). To a good approximation, the confinement free energy of a chain in a channel is equal to twice the confinement free energy of a chain confined to a slit [236, 239]

$$\mathcal{F} = (2/3)\pi^2\delta^{-2} \tag{5.11}$$

Additionally, extensive computational work on strongly confined wormlike chains has yielded an accurate prefactor to the Odijk expression for square channels [132]

$$\mathcal{F} = 2.2072\,\delta^{-2/3} \tag{5.12}$$

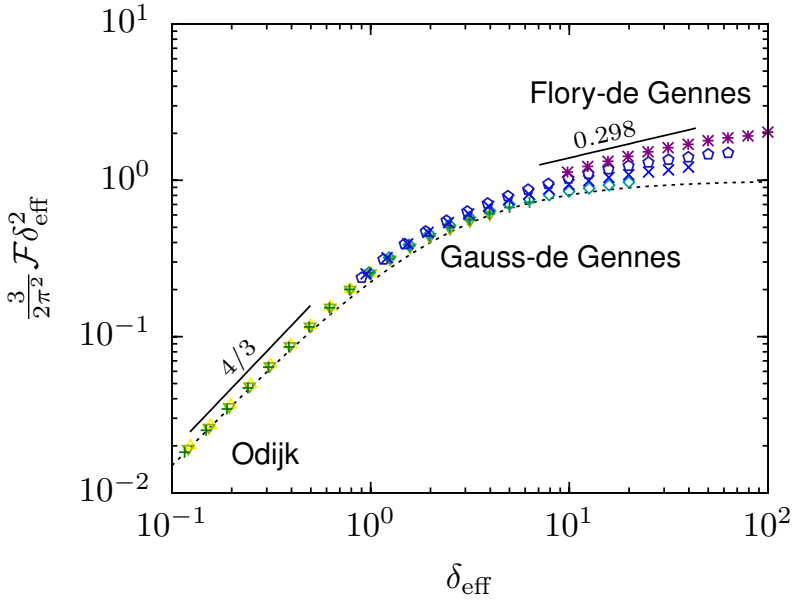

Figure 5.4: The normalized free energy of confinement as a function of normalized channel width, $\delta_{\rm eff} = (D - w)/l_p$ from Eq. 5.13 ($\epsilon = 0$, dashed line) and PERM simulations for $\epsilon = 2 \times 10^{-3}$ (open upward triangles), 5×10^{-3} (open downward triangles), 0.01 (plusses), 0.02 (open diamonds), 0.05 (crosses), 0.1 (open pentagons, DNA), and 0.2 (asterisks).

Following Chen and Sullivan [240] we propose an interpolation formula of the form

$$\mathcal{F} = \frac{(2/3)\pi^2\delta^{-2}}{(5.147\delta^{-2} + 3.343\delta^{-1} + 1)^{2/3}} \tag{5.13}$$

Taking the limit $\delta \to \infty$ yields Eq. (5.11) and $\delta \to 0$ matches Eq. (5.12). The remaining constant for the δ^{-1} term is used to fit the shape of the crossover region obtained from PERM simulations in the absence of excluded volume (see supporting information).

The similarity between Fig. 5.2 and Fig. 5.4 confirms the analogy between bulk and confinement, and the plateau in Fig. 5.4 validates the presence of a Gauss-de Gennes regime in confinement that connects the Odijk and Flory-de Gennes regimes. Compared to free solution, the Gauss-de Gennes regime in confinement is less prominent than the Gaussian regime in free solution because (i) the upper bound is lower in confinement (ϵ^{-1} versus ϵ^{-2}) and (ii) it is challenging to simulate extremely long chains with small ϵ at very high spatial resolution. Nevertheless, Fig. 5.4 clearly demonstrates the three

regimes, including the scaling exponent predicted by Eq. (5.9).

The close agreement here between the scaling theory and simulations has parallels with DNA confined in a sphere [241], but calls into question previous theories for the thermodynamics of a channel-confined chain between the Odijk and the Flory-de Gennes regimes. Most treatments apply Flory theory for a confined chain [136–138, 242] notwithstanding the fact that the accuracy of Flory theory predictions in free solution relies on a serendipitous cancellation of errors that are not *a priori* applicable in confinement. For example, the scaling $\mathcal{F} \sim \delta^{-4/3}$ predicted by Flory theory [30] for the "extended de Gennes" regime [71, 136–138] is not evident in our simulations. Other theories have attempted to incorporate backfolding of the chain to explain the transition between the Odijk and Flory-de Gennes regime [136, 140, 142, 237, 243]. The analogy between free solution and confinement makes the role of backfolding clear — it is simply the transition from rod-like to ideal statistics in the correlation volume.

5.4 Conclusion

Our results provide not only a complete description of the universal regimes of any long, channel-confined wormlike chain, but also have practical implications for genomic mapping in nanochannels [31, 47]. Our simulations predict that the Odijk regime is valid for an effective channel size $\delta_{\text{eff}} \equiv (D - w)/l_p \leq 0.3$ and the Flory-de Gennes regime begins at $\delta_{\text{eff}} \geq (2\epsilon)^{-1}$ (see supporting information). For DNA in a nanochannel, the Odijk extension [58, 132] applies for channels smaller than 20 nm, whereas the Flory-de Gennes extension [59] only starts to apply for channels larger than around 200 nm, where stretching is insubstantial. Since almost all experiments [30] and the commercial nanochannel technology [47] operate between these limiting cases, it is unsurprising that the experimental data are not described by the Odijk or de Gennes theories. Additionally, the Gauss-de Gennes regime certainly has implications for dynamics, which have recently been shown to be very sensitive to the anisotropy ϵ over similar ranges of extension (see Chapter 6). Future device design, as well as fundamental work, will need to account for the nature of the rod-to-coil transition of the subchains comprising nanoconfined polymers.

Chapter 6

Diffusion of DNA in a nanochannel

This chapter is based on the publication:
D. R. Tree, Y. Wang, and K. D. Dorfman
Phys. Rev. Lett. 108 (2012), 228105

6.1 Introduction

The configurations and dynamics of a flexible chain confined in a tube were described quite some time ago by de Gennes [59, 77, 130] and Odijk [58]. Emerging genomics technologies such as DNA barcoding [51, 56, 244, 245] have brought to the forefront the comparable problem of describing semiflexible chains when they are confined in a nanochannel [55, 60]. In this chapter, we show that the classical results for the mobility in the de Gennes [130] and Odijk regimes [58], which we will confirm describe the dynamics of flexible chains over the full range of confinement, are only the limiting cases for semiflexible chains such as DNA. Moreover, when DNA in a high ionic strength buffer is used as a model polymer, we predict that the mobility is independent of the fractional extension of the chain over the experimentally relevant range of chain extensions [60] ($\sim$ 20% to $\sim$ 80%). Thus, the commonly invoked ansatz [130] that the friction coefficient of a confined, semiflexible chain is proportional to its extension fails for DNA.

6.2 Theory

Let us first define what we mean by a semiflexible chain, since this term changes in different contexts [246]. The polymer is described by its contour length L, persistence length l_p, and effective width w, such that the chain consists of $N = L/l_p$ persistence lengths. Often, the term "semiflexible" is used in a global context to describe a chain where $L \approx l_p$, corresponding to a semiflexible filament such as actin. In our study of chains confined in nanochannels, we are concerned with the local flexibility of the chain on the length scale of the channel size, $D \approx l_p$. In this context the anisotropy of the "monomers" matters, with a flexible chain corresponding to $l_p/w \approx 1$ and a semiflexible chain corresponding to $l_p/w \gg 1$ [72, 210].

In particular, we will focus on double-stranded DNA in a high ionic strength buffer that screens electrostatic interactions, which has frequently been used as a model system for a confined polymer [109]. In these conditions, DNA is clearly a semiflexible chain, with $l_p = 53$ nm [64] and $w = 4.6$ nm [136]. As we will see, this high degree of anisotropy limits de Gennes' model [59, 77, 130] to very small values of the fractional extension. The DNA used in experiments can be quite long, normally tens of microns in length. As a result, the chain is flexible in the global sense since $L \gg l_p$.

We already know that the semiflexible nature of DNA strongly affects its equilibrium extension [71, 136, 138, 139, 238]. Figure 6.1 shows how the average chain extension, $\langle X \rangle$, depends on the degree of confinement for a flexible chain and semiflexible chain. These data were generated by modeling the chain as a series of $N_b = 2048$ touching beads [122] of size w that interact by a hardcore excluded volume interactions. To give the chain a persistence length of l_p, a bending potential is enforced between trios of beads according to the discrete wormlike chain model [71, 190]. Analogous to our prior work [71], we generated an equilibrium ensemble of chain configurations using Monte Carlo simulations with reptation, crankshaft and pivot moves[1] . The simulation was run in each case until the statistical errors, corrected for the time series autocorrelation [247], were smaller than the size of the plot symbols.

The classical theories [58, 59, 77] provide a complete description for the extension

[1] For the persistence lengths of (5.3, 23, 53) nm, we obtained (10^3, 2×10^2, 10^3) samples per simulation using (48, 48, 12) independent simulations, each with an equilibration of (2.07, 8.19, 20.6) $\times 10^8$ steps and production runs of (2.05, 2.05, 4.10) $\times 10^9$ steps.

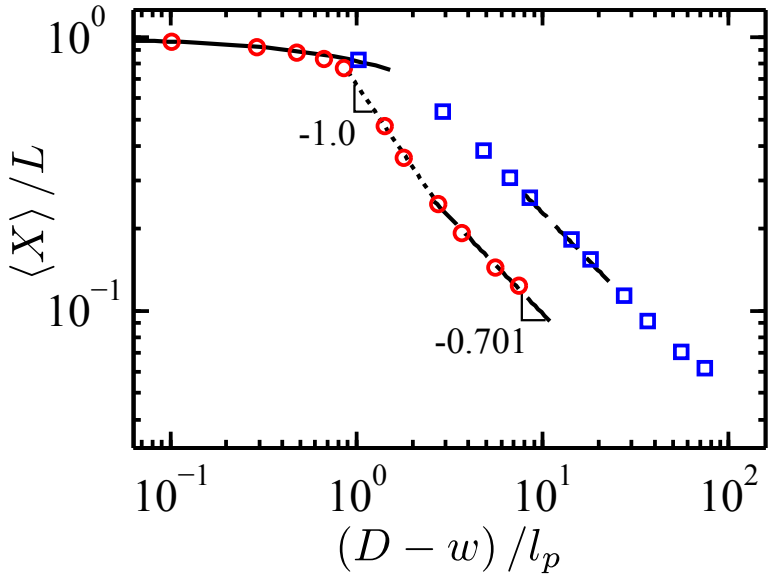

Figure 6.1: Averaged extension of a flexible ($l_p = 5.3$ nm, squares) and a semiflexible ($l_p = 53$ nm, circles) chain containing 2048 touching beads of width $w = 4.6$ nm as a function of the effective channel width, $D - w$, available to the chain. To aid the eye, lines corresponding to the Odijk regime (solid), transition regime (dotted), and extended de Gennes/de Gennes regimes (long dashed) are shown.

of the flexible chain. Over almost the full range of extension, the flexible chain is in the de Gennes regime [59, 77]. Here, the chain consists of isometric compression blobs of characteristic volume D^3 containing a subchain of length $L_{\rm sub} \cong D^{5/3}(wl_p)^{-1/3}$ [138]. The corresponding extension is $\langle X \rangle \cong L(wl_p)^{1/3}D^{-2/3}$. A more precise calculation yields $\langle X \rangle \sim D^{(\nu-1)/\nu}$ with $\nu = 0.5877$ being the Flory parameter [71]. In the tightest channels, the chain crosses over into the Odijk regime [58], where the chain consists of a series of deflection segments. The extension here is $\langle X \rangle = L[1 - 2\alpha\,(D/l_p)^{2/3}]$ with $\alpha = 0.09137$ a universal prefactor [131, 132].

In contrast, we already know [71, 136, 138, 139, 238] that the classical theories [58, 59, 77] only correspond to the limiting cases for the extension of a semiflexible chain. Indeed, in order for a semiflexble chain to be able to reach a de Gennes regime, the polymer must have a length of at least $L \cong {l_p}^3/w^2$ in a channel that is larger than $D \cong {l_p}^2/w$ [71, 138]. As a semiflexible chain is compressed by decreasing the channel size, the blobs become anisometric [71, 136, 138] with size D^2H, where $H \cong (Dl_p)^{2/3}w^{-1/3}$. Each one of these cylindrical blobs contains a subchain of length $L_* \cong {l_p}^{1/3}D^{4/3}w^{-2/3}$. This regime was named the "extended" de Gennes regime [71] because the scaling for the extension in the de Gennes regime, $\langle X \rangle \cong L(wl_p)^{1/3}D^{-2/3}$, extends to the case of anisometric compression blobs. When the channel size approaches the order of the persistence length, $D \approx l_p$, the chain can no longer form blobs. Here the behavior crosses into a transition regime where several simulations [71, 139, 238], as well as our results in Fig. 6.1, indicate that the extension scales like $\langle X \rangle \sim D^{-1}$ [71, 139, 238]. The free energy of these configurations is unknown, and it is not clear yet if the behavior is universal. Finally, when $D \ll l_p$, the other classical limit of Odijk [58] is recovered.

For DNA confined in a nanochannel, semiflexibility is a crucial aspect. As the anisotropy of the monomers increases, the width of the transition regime grows; the maximum extension in the extended de Gennes regime is compressed to $\langle X \rangle / L \cong (w/l_p)^{1/3}$. When DNA in a high ionic strength buffer is used as a model for a confined polymer [55, 60], the extended de Gennes regime and, in particular, the transition regime encompass almost the entire experimental range of extensions [71]. Indeed, the existence of these additional regimes explains [71] the disagreement between early experiments on DNA extension in nanochannels [60] and the de Gennes model.

Let us now consider the mobility of a confined semiflexible chain. By applying an

infinitesimal force f_x that is uniformly distributed along the chain, the corresponding velocity along the channel axis is

$$v_x = \mu f_x = \langle \Omega_{xx} \rangle f_x \tag{6.1}$$

where μ is the mobility of the chain. As seen in eq. (6.1), we can obtain the Kirkwood approximation to the mobility [116, 118] from the appropriate component of the hydrodynamic tensor, Ω_{xx}, where the brackets refer to an average over the equilibrium distribution of chain configurations.

For a flexible chain, the number of monomers inside the volume D^3 where the walls screen long-range hydrodynamic interactions is sufficiently high to permit a simple scaling law. Simplifying eq. (6.1) in terms of the pair-correlation function, $g(r)$, following de Gennes [77], yields

$$\mu = N^{-1} \int g(r)\Omega(r) d^3\mathbf{r} \tag{6.2}$$

In the blob theory [130] the pair-correlation function is replaced with c, the number concentration of segments inside a blob, and the hydrodynamic screening by the walls is approximated by $\Omega(r) = 1/\eta r$ for $r < D$ and an exponential decay for $r > D$ [77, 130], where η is the solvent viscosity. Since we only need an approximate result, the remainder of the calculation is simplified by using spherical coordinates and integrating over the solid angle [77],

$$\mu = \frac{4\pi c}{N} \int_0^D \frac{1}{\eta r} r^2 dr \approx \frac{cD^2}{\eta N} \tag{6.3}$$

In the de Gennes regime, the monomer concentration in the blobs is $c \cong (L_{\mathrm{sub}}/l_p)/D^3$, which yields $c \cong w^{-1/3} l_p^{-4/3} D^{-4/3}$. Recalling that $N = L/l_p$, we recover the classic result [130]

$$\mu \sim (1/\eta L) \langle X/L \rangle^{-1} \tag{6.4}$$

In the extended de Gennes regime, the density of segments is $(L_*/l_p)/(D^2 H)$, which again yields $c \cong w^{-1/3} l_p^{-4/3} D^{-4/3}$. As a result, the blob theory predicts the diffusion in the extended de Gennes regime is also given by eq. (6.4).

The key assumption leading to eq. (6.4) is that the number of segments in the screening volume, cD^3, is large enough so that each blob is non-draining (Zimm). In other words, the subchain comprising a blob entrains the fluid inside it, whereupon the segment-segment hydrodynamics dominate and the subchain behaves hydrodynamically

like a solid object. Free draining (Rouse) behavior at the subchain level should arise when $D \approx 2l_p$. There is now approximately one Kuhn length inside D^3, which causes each segment to be an independent friction center. In other words the segment-fluid hydrodynamic interactions are dominant. In this limit, we would expect

$$\mu \sim (1/\eta L) \langle X/L \rangle^0 \tag{6.5}$$

The question is whether the chain reaches the scaling of eq. (6.5) before it reaches the Odijk regime ($D \ll l_p$). In the latter case, the chain is like a slender, confined rod. Its mobility [248]

$$\mu \approx \frac{1}{2\pi} \ln \left(\frac{l_p}{2a} \left[\frac{1 - \langle X/L \rangle}{2\alpha} \right]^{3/2} \right) \tag{6.6}$$

reflects the dominance of segment-wall interactions. The latter expression involves the bead hydrodynamic radius, a. We chose $a = 1.38$ nm so that the chain mobilities in free solution for $l_p = 53$ nm matched experimental values for DNA [92, 108, 193]. While we raise this issue for nanochannels, similar concerns about the draining behavior have been expressed for DNA in slits [152, 154].

6.3 Results and Discussion

To determine if and when the chain crosses over to eq. (6.5), we computed the Kirkwood mobility through a Monte Carlo integration of eq. (6.1) [90, 188, 203, 249]. For a given chain configuration, we computed the 3×3 chain hydrodynamic tensor

$$\mathbf{\Omega} = \frac{1}{{N_b}^2} \sum_{i,j}^{N_b} \left[\frac{\delta_{ij}}{6\pi\eta a} \mathbf{I} + (1 - \delta_{ij}) \mathbf{\Omega}^{\mathrm{OB}}(\mathbf{r}_{ij}) + \mathbf{\Omega}^{\mathrm{W}}(\mathbf{r}_i, \mathbf{r}_j) \right] \tag{6.7}$$

In the latter, δ_{ij} is the Kronecker delta, $\mathbf{r}_i$ and $\mathbf{r}_j$ are the positions of bead i and j respectively and $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$. The hydrodynamic tensor includes a self-diffusion term, a free-solution Oseen-Burgers tensor [176], $\mathbf{\Omega}^{\mathrm{OB}}$, and a wall term, $\mathbf{\Omega}^{\mathrm{W}}$, due to the effects of the no-slip condition at the channel boundaries. The Oseen-Burgers tensor is acceptable in this calculation because the beads are hard spheres, and do not suffer from unphysical behavior caused by bead-bead overlap. The wall term was calculated using a numerical solution of Stokes equation, similar to Jendrejack *et al.* [176]. We employed

a second-order finite difference approach with a staggered, three-dimensional, uniform, Cartesian mesh [177] and mass-conserving boundary conditions. Due to the prohibitive computational time needed to solve the hydrodynamic problem for each chain configuration, the wall term was calculated and stored on a grid, and subsequently linearly interpolated during Monte Carlo averaging. Finally, we note that in each case the statistical errors of the computed diffusivity, corrected for the time series autocorrelation [247], are smaller than the size of all plot symbols.

Figure 6.2a shows the results for the mobility of DNA as a function of its extension. In the largest channels, corresponding to the smallest fractional extensions, the channel provides minimal confinement and the chains are approaching the Zimm free solution mobility, $\mu \sim L^{-3/5}$. Outside of this limit, the friction due to the walls is substantial. If we neglect the wall term in eq. (6.7) for a channel size of 80 nm, the resultant mobility is more than 5 times larger.

The key result is that the Rouse scaling in eq. (6.5) encompasses all of the extensions seen in experiments for DNA [60]. In contrast, Fig. 6.2b shows that eq. (6.4) is a reasonable description for the flexible chain all the way to the transition to the Odijk regime of eq. (6.6). To be more quantitative, linear regression gives $\mu \sim \langle X \rangle^{-0.874}$ ($R^2 = 0.998$) which agrees very well with the value of $\mu \sim \langle X \rangle^{-0.61/0.7015}$ obtained from other flexible chain calculations [250]. We also simulated an intermediate persistence length $l_p = 23$ nm and found an intermediate result; for short extension this chain obeys de Gennes scaling but it still exhibits a broad transition towards the Odijk result.

In the case of the flexible chain, the crossover between de Gennes and Odijk mobilities is narrow, mirroring the extension behavior. If the confinement does not force a rod-like conformation, this chain is so flexible that it can only form non-draining blobs. In the semiflexible case, the large monomer anisotropy leads to a wide gap between the de Gennes regime and the Odijk regime for the extension. This gap closely aligns with the beginning and ending of the mobility plateau in Fig. (6.2)b. Thus the existence of additional regimes for the extension of semiflexible chains explains both the existence of the mobility plateau and the fact that it grows with increasing persistence length.

While we have focused exclusively on the dynamics of DNA in a high ionic strength buffer, where electrostatic interactions are screened, there are DNA barcoding devices [51, 244] that use high ionic strengths to stiffen the DNA backbone. As the ionic

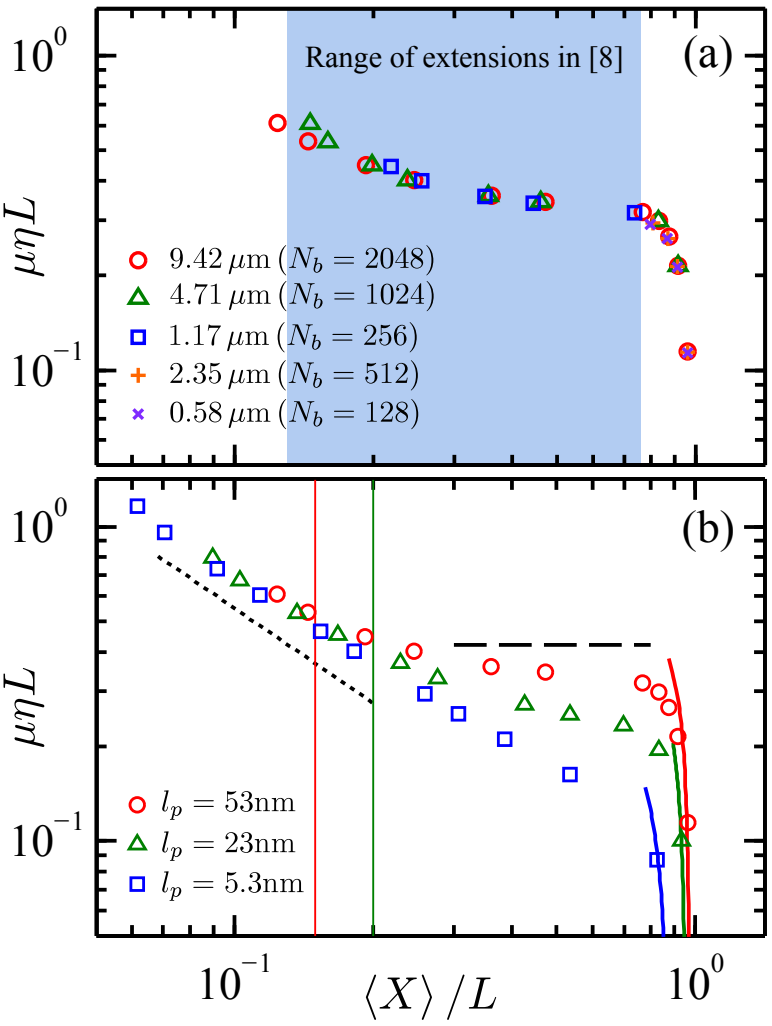

Figure 6.2: Mobility versus extension. All simulations correspond to $w = 4.6$ nm and $a = 1.38$ nm. (a) Results for five different chain lengths for $l_p = 53$ nm. The shaded region corresponds to the extensions seen in DNA experiments [60]. (b) Results for three different persistence lengths for $L = 9.42$ μm ($N_b = 2048$ beads). The dotted line is the scaling of eq. (6.4) and the dashed line shows the scaling of eq. (6.5). The solid lines are the approximation in eq. (6.6). The vertical lines are the values for the onset of the scaling $\langle X \rangle / L \sim D^{-1}$ for the 53 nm chain ($\langle X \rangle / L = 0.15$) and the 23 nm chain ($\langle X \rangle / L = 0.2$).

strength decreases, the predicted values for the effective width and persistence length begin to converge [66]. Our analysis thus predicts that DNA will obey the de Gennes prediction in eq. (6.4) in a sufficiently low ionic strength such that $l_p/w \approx 1$ and a large enough channel such that this very high persistence length chain can form compression blobs. These experiments are technically challenging, since the length of DNA required to reach the de Gennes regime in a low ionic strength buffer is enormous.

6.4 Conclusion

In this chapter, we have clearly shown that the hydrodynamics of confined semiflexible chains deviate significantly from the classic prediction for a flexible chain in eq. (6.4) [77, 130]. As there are a large number of publications using DNA in a high ionic strength buffer as a model polymer, it is important to keep in mind the stark differences between the dynamics of semiflexible chain such as DNA and the more flexible chains often encountered in polymer physics [109].

Chapter 7

Relaxation Time of DNA in Nanochannels

This chapter is based on the publication:
D. R. Tree, Y. Wang, and K. D. Dorfman
Biomicrofluidics 7 (2013), 054118

7.1 Introduction

The extension of a long DNA molecule confined in a nanochannel has attracted tremendous attention [28–31] in large part because it couples a fundamental polymer physics problem to an important application in genomics, namely DNA barcoding [10, 47, 51, 56, 251]. The theoretical basis for the equilibrium extension of DNA in a nanochannel has been addressed to varying degrees of accuracy by theory and simulation for strong [51, 58, 131, 132, 142, 190, 244, 246, 252, 253], moderate [71, 136, 138–140, 142, 155, 237, 238, 243, 254–257] and relatively weak [59, 176, 258, 259] confinement, leading to the reconciliation [71] between early experimental observations [60] and the predictions of the classic theories from Odijk [58] and de Gennes [59] (see also Chapter 5). Due to both the paucity of dynamic data in strong confinement and the computational difficulty of simulating the dynamics of long chains, there is little work validating the computational predictions of confined DNA dynamics in nanochannels with hydrodynamic interactions (see Chapter 6). We show here that the confined wormlike chain model used to model

DNA extension can explain experimental results [55, 60] for the relaxation time of a long DNA molecule in a nanochannel. Furthermore, our simulations highlight a previously overlooked strong reduction in the relaxation time as the channel size is reduced below the Kuhn length of the DNA, which has practical implications for the practice of DNA barcoding [47].

It is not trivial to compute the relaxation time of a long polymer in confinement. In addition to obvious challenges in computing the hydrodynamics of confined polymers [260], one of the major challenges in modeling DNA in a nanochannel is selecting a model that can accommodate the large separation of length and time scales. Explicitly, the effective width of the polymer backbone, w, is small compared to the persistence length, l_p, which itself is small compared to the contour length, L, of the chain. The nanochannel width introduces a fourth length scale, D, with the ratio D/l_p determining the strength of the confinement. In very weak confinement ($D \gg l_p$) it is possible to use a coarse-grained, bead-spring model [176] parameterized to match experimental data [141, 191]. However, this model cannot resolve DNA deformation when the confinement length scale approaches the undeformed size of one (Gaussian) "spring" of DNA. A natural solution to the resolution problem is to use a fine-scale model, such as the wormlike osculating-sphere model (WOSM) described in Chapters 5 and 6 and Refs. 75, 122, where the DNA is modeled by a string of beads of size w interacting via a discrete wormlike chain bending potential [190] and hard core excluded volume. This model can easily capture confinement down to the strong confinement regime ($D \ll l_p$), provided that we simultaneously ensure that the bond length, which is the bead size w for the WOSM, is small compared to the channel size to avoid discretization artifacts. These are not the only two model options; for example, the discrete stretchable, shearable wormlike chain model [261] represents an attractive choice, but the dynamical implementation of this model for complicated problems remains a work in progress. There are other options for simulating confined polymers that are well suited to examining scaling laws, such as the bond fluctuation method [262], but these models are challenging to connect quantitatively to the experimental parameters.

For double-stranded DNA in nanochannels in a high ionic strength buffer, it is now accepted that most experiments are carried out in the crossover regime between the classical de Gennes regime [59] (suitable for $l_p^2/w < D < R_g$ [71, 136], where

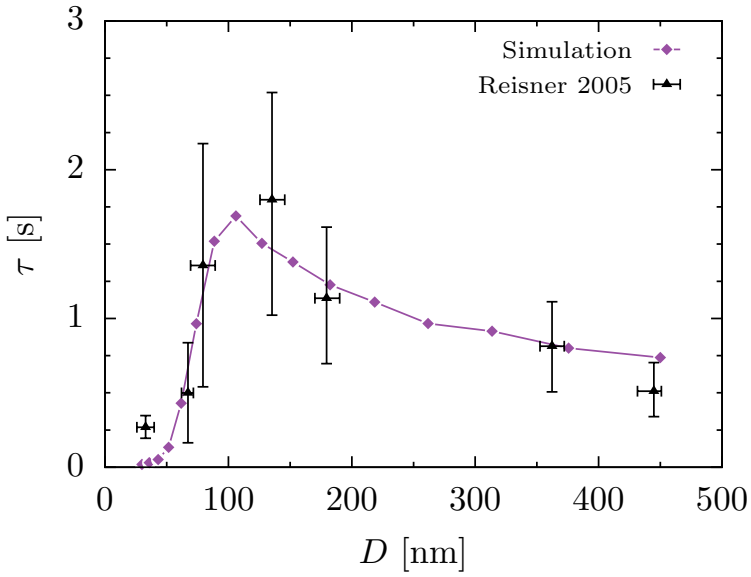

Figure 7.1: Relaxation time (diamonds) obtained from Eq. 7.5 with $c = 1.2$ and the WOSM compared to the experimental data of Reisner *et al.* [60] (triangles with error bars).

R_g is polymer's radius of gyration in free solution) and Odijk regime [58] (suitable for $D \ll l_p$). These circumstances necessitate a fine-scale representation such as the WOSM. However, a typical molecule such as λ-DNA (48,500 base pairs) requires simulation of several thousand beads and the longer molecules used for genomic mapping [47] require tens of thousands of beads. Dynamic simulations become extremely expensive with high spatial discretization and large molecular weights, even with fast implicit solvent methods for the hydrodynamic interactions in confinement [260].

Fortunately, it is possible to arrive at a reasonable estimate for the polymer relaxation time, τ, by mapping the chain dynamics to a one-dimensional, overdamped, Rouse dumbbell model with a finite but non-zero equilibrium extension [118, 263]. In the present contribution, we show how the Monte Carlo methods used in our previous work to compute the extension in Chapter 5 and Ref. 71 and the hydrodynamic mobility in Chapter 6 can be used to parameterize such a dumbbell model. The ultimate result of our analysis is seen in Fig. 7.1, which compares our computational predictions to the seminal experiments of Reisner *et al.* [60]. The agreement between our approach and

the experimental results shown in Fig. 7.1 lend confidence to our use of the WOSM, the Kirkwood approximation employed to obtain the friction, and the assumptions made in mapping the model to an overdamped, Rouse dumbbell. Furthermore, this agreement provides a basis for considering the WOSM model in the engineering of nanochannel devices for genomic mapping. However, there are certainly limitations to the computational model and gaps in our understanding of the underlying physics, which are discussed in conjunction with our modeling results.

7.2 Methods

7.2.1 Dumbbell Model and Mapping

We begin by recalling the physics required to map the chain dynamics to a one-dimensional, overdamped, Rouse dumbbell model with a finite equilibrium extension [118, 263]. In this model, the autocorrelation function $C(t)$ of the fluctuation about the mean extension $\langle X \rangle$ is given by[264]

$$C(t) = \langle \delta_X(0)\delta_X(t) \rangle = \langle \delta_X^2 \rangle \exp(-t/\tau) \tag{7.1}$$

where

$$\delta_X(t) \equiv X(t) - \langle X \rangle \tag{7.2}$$

is the deviation from the mean extension and

$$\tau = \frac{\zeta}{2k_{\text{eff}}} \tag{7.3}$$

is the relaxation time. Two important terms appear in Eq. 7.3: ζ, the friction coefficient of each of the two beads comprising the dumbbell and k_{eff}, the spring constant of the Hookean spring between them [118, 263].

To map between the two models, both the friction coefficient, ζ, and the effective spring constant, k_{eff}, of the dumbbell model must be defined in terms of the wormlike chain model. Since the dumbbell model has no hydrodynamic interactions, the friction of the center-of-mass of the dumbbell is simply equal to 2ζ. Equating the center-of-mass friction of the WOSM to that of the dumbbell defines ζ and provides the first part of the map. Per this definition, the dumbbell friction is independent of conformation

fluctuations. This is appropriate, since the friction coefficient obtained from the WOSM employs a rigid-body approximation [124, 203] similar to the mean field pre-averaging approximation used in the Zimm model [118]. Accordingly, we focus on the case when the conformation is unperturbed by external forces (other than confinement). This means the relaxation time obtained by this method is only valid for the fluctuations of the polymer about its equilibrium conformation. We thus do not consider the possibility of a second relaxation time related to non-equilibrium stretching [156].

The spring constant of the Rouse dumbbell is obtained from the equipartition theorem, which gives

$$k_{\text{eff}} = \frac{k_B T}{\langle \delta_X^2 \rangle} \tag{7.4}$$

where k_B is Boltzmann's constant, T is the absolute temperature and $\langle \delta_X^2 \rangle$ is the variance in the extension of the spring. Since the spring is harmonic, the probability density function of the extension in the dumbbell model is Gaussian [118], and is therefore completely described by the mean span, $\langle X \rangle$, and variance $\langle \delta_X^2 \rangle$. Subsequently, the first two moments of the extension distribution function in the WOSM are used to define the dumbbell harmonic spring and are thus sufficient to define the effective spring constant. This implies that the spring constant mapping is only valid insofar as the extension distribution function is well described by its first two moments. This certainly breaks down for extensions near the maximum contour length and cases where the extension distribution is complicated or multi-modal as might exist in the presence of backfolded states in tight confinement [140, 142, 254]. Note that for the chain extension we chose to work in terms of the span, X [71], rather than the end-to-end distance, since the former is the experimentally relevant metric.

In the light of this mapping, it proves convenient to recast Eq. 7.3 in terms of the quantities obtainable from the simulation of the WOSM. With this change of variables, the relaxation time becomes

$$\frac{\tau}{\tau_{\text{R}}} = \frac{c}{4} \frac{\langle \delta_X^2 \rangle}{L l_p} \frac{2\zeta}{\eta L} \tag{7.5}$$

where

$$\tau_{\text{R}} \equiv \frac{\eta L^2 l_p}{k_B T} \tag{7.6}$$

is the Rouse relaxation time for an ideal chain without the prefactor [72, 114]. In Eq. 7.5, we included a prefactor c as an $O(1)$ fitting parameter that we will use in the subsequent

analysis. Note that $c = 4/\pi^2$ would be an appropriate choice resulting from matching the center of mass diffusivity of a Rouse chain to a dumbbell model [264].

While a simple dumbbell model may seem like a very crude approximation, a single exponential decay of $C(t)$ is consistent with experimental findings [30, 60]. Furthermore, we do not construct the dumbbell arbitrarily; instead, it is parameterized to match equilibrium properties determined from Monte Carlo sampling of the fine-scale WOSM. This strategy gives nearly quantitative agreement between the simulation results and experimental data for the relaxation time with a single $O(1)$ fitting parameter to account for the uncertainty in the various physical parameters appearing in the detailed model.

7.2.2 Monte Carlo Simulations

To parameterize the dumbbell model, we employ the aforementioned WOSM in a square nanochannel of size D as described in Chapters 4–6. To compute equilibrium chain properties, we used a standard Metropolis algorithm with reptation and crankshaft moves [71]. To calculate the values of ζ and $\langle \delta_X^2 \rangle$ which we desire, we require a fully parameterized WOSM, which includes an effective width, w, a persistence length, l_p, a contour length, L, and a hydrodynamic radius, a (see Chapter 4. As the aim of our simulations is ultimately a comparison with the experimental relaxation time data from Reisner *et al.* [60], we examine λ-DNA in channel sizes between 30 nm and 450 nm. We approximate the dyed contour length of λ-DNA as $L = 18.63$ μm, as suggested by these authors [60], which is within the sizes we can compute using the standard Metropolis algorithm (see Chapters 3–5 and Ref. 71).

In addition to the channel size and contour length, the model requires specifying the persistence length, which parameterizes the bending energy, and an effective width of the DNA, which parameterizes both the hard core, intrachain excluded volume interactions and the hardcore interactions with the channel walls. Note that in the WOSM the discretization length equals the effective width making the total contour length equal to the number of beads times the effective width. While there is widespread agreement that the persistence length of double-stranded DNA in a high ionic strength buffer is approximately 50 nm [64], there is less consensus surrounding the effective width.

In our previous work [71], we suggested treating the effective width as a free parameter to fit the simulation data for $\langle X \rangle$ to experimental data, provided that the end

result is reasonably close to the prediction from Stigter's theory for short, rod-like DNA [219]. (For high ionic strengths, Stigter's theory[219] predicts $w \approx 5$ nm.) In our initial foray into this problem,[71] we concluded that an effective width between 4.6 nm and 12 nm seemed reasonable, with $w = 7$ nm being the best fit for the experimental extension data [60] after trying to collapse it with de Gennes scaling. Given the uncertainty in the exact value of the effective width, especially the difference between DNA-DNA interactions and DNA-wall interactions, we decided to take a simple, approximate approach in our work here. We set the effective width to be $w = 10$ nm and combine this with an estimate of the persistence length with a single significant digit as well, $l_p = 50$ nm. These are reasonable order-of-magnitude estimates for both parameters, and we conjecture that the adjustable parameter c in Eq. 7.5 will ultimately allow us to correct for the uncertainty.

Figure 7.2(a) shows the WOSM mean span versus channel size compared to the experimental data of Reisner *et al.* [60], which justifies our supposition regarding the parameter estimates. There is very good agreement between the model predictions and the experimental data for large channel sizes, where we observe de Gennes scaling and expect excluded volume to play a significant role (see Chapter 5). However, for small channels, the WOSM values deviate significantly from the experimental data, suggesting that as the channel size approaches the effective width, hardcore interactions between the polymer and channel become increasingly inadequate to describe the real system. We should also point out that the experimental data were obtained in rectangular channels [60], and there is a small correction for the channel aspect ratio [71] that is not incorporated into our analysis.

Figure 7.2(b) shows that the extension results are also consistent with the scaling produced by prior simulation work. In the region $D/l_p \sim 1$, we observe a slope D^{-1} similar to the data in Chapter 5 and to simulation data by other groups [71, 139, 238]. This slope is also consistent with the existence of a Gauss-de Gennes regime, as discussed in Chapter 5. For the larger channel sizes, the scaling for the extension switches to $D^{1-1/\nu}$ with $\nu = 0.5877$. The latter result is consistent with either a de Gennes regime [59] or an extended de Gennes regime [71, 138].

In addition to equilibrium values, we are interested in the chain friction, ζ, which is determined by a rigid-body approximation to the diffusivity [122, 249]. Specifically,

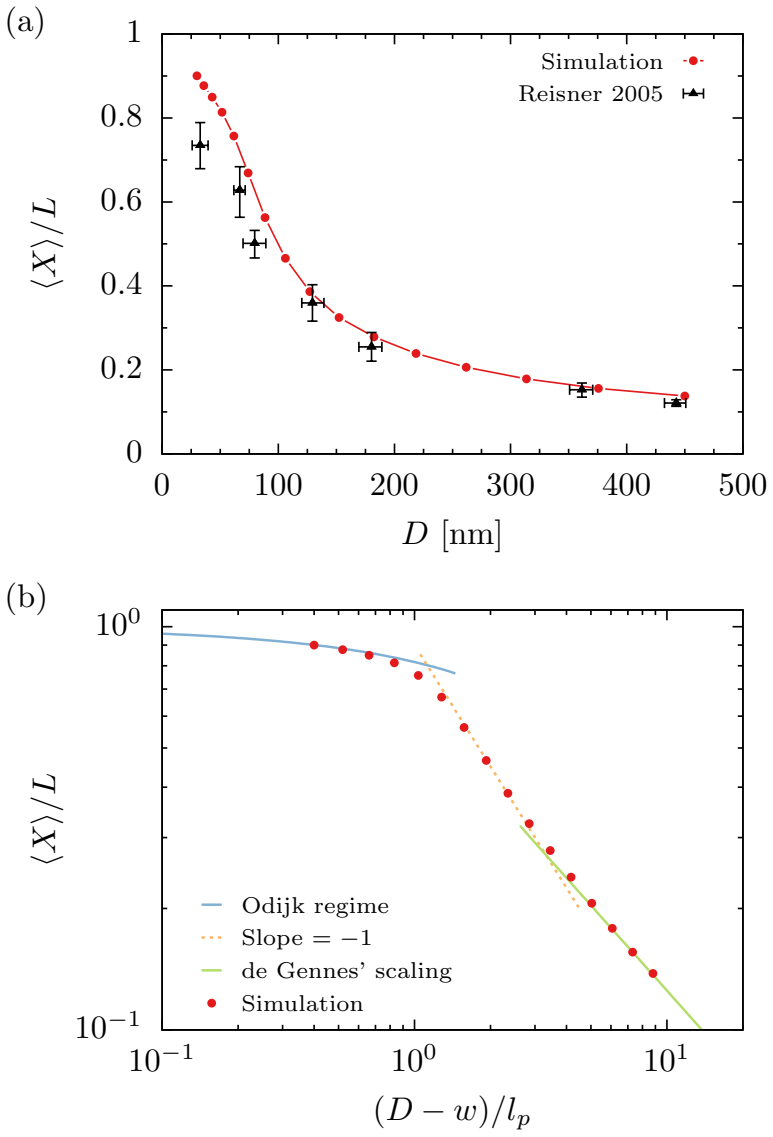

Figure 7.2: (a) Fractional extension of DNA in a square nanochannel of size D from Monte Carlo simulations (circles) and from Reisner *et al.* [60] (triangles). Error bars for the simulation data are the standard error, and when not explicitly shown, the error is of order of the symbol size. Simulation parameters: $l_p = 50$ nm, $w = 10$ nm, $L = 18.63$ μm (i.e. 1863 spheres). The same set of parameters is used throughout this work. (b) Same data plotted in dimensionless log-log form. The solid curve is the prediction for the Odijk regime [58] with no free parameters [132]. The dashed line is the predicted slope for the proposed Gauss-de Gennes regime (Chapter 5), and the solid line in the bottom right is the scaling $D^{1-1/\nu}$ with $\nu = 0.5876$.

similar to Chapter 6, the chain conformations used to obtain Fig. 7.2 are combined with a numerically determined confined hydrodynamics tensor [176]. The confined hydrodynamic tensor leaves a degree of freedom for the bead hydrodynamic radius, a. Our analysis of the free solution diffusivity of DNA in Chapter 4 indicates that for a touching *hydrodynamic* bead model, a value of $a = 3$ nm gives a good approximation of the diffusivity. We use this value, noting that the WOSM model does not have touching hydrodynamic beads, which introduces some error. Doing so, we again anticipate that the value of c in Eq. 7.5 will compensate for the error introduced here.

Before proceeding, we should also justify the need for the new Monte Carlo simulations presented here in the light of our previous work. In principle, we could use our existing simulation data for short chains from Chapters 5–6 and Ref. 71, and then extrapolate to λ-DNA using scaling laws. However, since these scaling laws are implicitly one of the things we are trying to evaluate, such an approach seems questionable. The new calculations we used here are intended to provide data at the molecular weight of λ-DNA and cover the full range of channel sizes used in experiments [60]. We thus expect the trends in the mean span [71] and hydrodynamic mobility as a function of confinement to be identical to our prior results, even if the quantitative values may differ slightly. For instance, the parameters used in Fig. 7.2 correspond to a monomer anisotropy $\epsilon \equiv w/l_p = 0.2$. In the discussion in Chapter 5 on the Gauss-de Gennes regime, we pointed out that the Gauss-de Gennes scaling $\langle X \rangle \sim D^{-1}$ results from the stiffness of the chain, much in the way that the range of molecular weights that exhibit ideal chain scaling for the free solution radius of gyration increases with monomer anisotropy. Using scaling arguments, we showed that the range of channel sizes in the Gauss-de Gennes regime increases with increasing monomer anisotropy, and that this regime should disappear entirely in the limit of a freely jointed chain. The relatively flexible chain model used here suppresses the extent of the Gauss-de Gennes scaling $\langle X \rangle \sim D^{-1}$, which is reflected in the data in Fig. 7.2.

Additionally, we have also computed new results for the variance in the extension (span) of the chain, a parameter appearing in the dumbbell model, and (in Section 7.3.2) we confirm that the span distributions are reasonably Gaussian, which is an assumption in the dumbbell model. The variance of the span has yet to be systematically studied as a function of confinement and monomer anisotropy, although there are some intriguing

new results supporting the existence of an extended de Gennes regime [155]. In fact, all of the physical quantities required to arrive at the relaxation time (average extension, variance in the extension, and the hydrodynamic mobility) depend on the extent of confinement and the physical properties of DNA, and a complete understanding of these relationships has yet to be achieved [30]. Thus, while there are deep scientific insights to be mined by exploring these dependencies in detail, for this work we keep our focus very practical — we simply want to show that models and methodologies now exist to determine the relaxation time of channel-confined DNA, and carefully assess the limitations of this approach.

7.3 Results and Discussion

7.3.1 Computing the relaxation time

Figure 7.3 shows Metropolis Monte Carlo results for the variance in the extension, $\langle \delta_X^2 \rangle$, and the chain friction, ζ, as a function of channel size for λ-DNA. The friction and mean span fluctuations in Fig. 7.3 along with the parameters necessary to define τ_R provide the necessary data to compute the relaxation time via Eq. 7.5. Using the aforementioned contour and persistence lengths and assuming a viscosity of 1 cP and a temperature of 298 K gives $\tau_R \approx 4.2$ s. Combining these results with the choice $c = 1.2$ gives the relaxation times shown in Fig. 7.1, which are compared to the experimental data reported by Reisner *et al.* [60].

The friction coefficient appearing in Fig. 7.3 behaves as expected, given previous computational results for the mobility of semiflexible chains in confinement (Chapter 6). In weak confinement, the friction slowly increases as the channel size decreases, consistent with a blob theory. In strong confinement, the friction diverges logarithmically, consistent with a lubrication model for the relative motion of concentric cylinders. From this we conclude that these previously observed trends are robust to the molecular weight, persistence length and effective width used here.

The variance, on the other hand, shows more interesting behavior. The data show a small, gradual increase in the variance as the channel size decreases until the channel size is around the Kuhn length of DNA (100 nm). Although we are focusing primarily on the approach required to produce Fig. 7.1, it is worthwhile to make a brief diversion

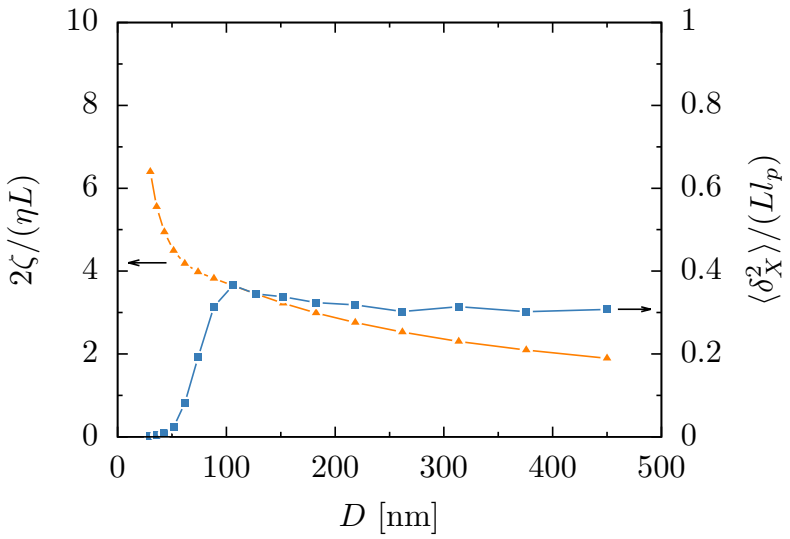

Figure 7.3: Normalized chain friction (left panel, triangles) and normalized fluctuations in the mean span (right panel, squares) of the WOSM as a function of channel size.

here to discuss the relatively flat response of the channel fluctuations for $D > 100$ nm. In a recent publication, Dai and Doyle [155] proposed that the fluctuations in the chain extension are given by

$$\langle \delta X^2 \rangle \cong N_{\text{blob}} R_{\text{blob}}^2 \tag{7.7}$$

where N_{blob} is the number of blobs and R_{blob} is the size of the blob. The ideas behind this equation are that (i) the fluctuation of each blob is independent of the other blobs and (ii) the influence of confinement inside a blob is negligible. Their analysis then suggests that the fluctuations in the extended de Gennes regime [138] should have the scaling [155]

$$\langle \delta X^2 \rangle \cong L l_p \tag{7.8}$$

which agrees with the result obtained from Flory theory [71] for the extended de Gennes regime. Interestingly, this logic also leads to the same scaling for the Gauss-de Gennes regime introduced in Chapter 5, if blob theory is indeed valid there. In the latter regime, the blob size is the channel size,

$$R_{\text{blob}} \cong D \tag{7.9}$$

The blobs are assumed to exhibit ideal chain statistics [59, 255, 265]

$$L_{\text{sub}} l_p \cong D^2 \tag{7.10}$$

where L_{sub} is the contour length of the subchain inside a blob. The total number of blobs is then

$$N_{\text{blob}} = \frac{L}{L_{\text{sub}}} \cong \frac{L l_p}{D^2} \tag{7.11}$$

Using Eq. 7.9 and Eq. 7.11 in Eq. 7.7 yields the same scaling for the extension fluctuations in the extended de Gennes regime (Eq. 7.8) and the Gauss-de Gennes regime.

Regardless of the specifics of the regime, the lack of dependence of the variance on the relaxation time means that, in moderate confinement, the relaxation time is especially sensitive to the friction coefficient. As a result, the qualitative agreement between the calculated and experimental relaxation times appearing in Fig. 7.1 provides evidence that, in moderate confinement, the rigid-body assumptions used to obtain the friction coefficient are adequate. While this evidence is not as strong as a direct experimental measurement of the chain friction coefficient, it is still a significant result given the difficulty in obtaining both experimental and computational data for the dynamics of moderately confined semiflexible chains.

As the channel size is decreased further and falls below the Kuhn length, the variance quickly drops, presumably due to the loss of the degrees of freedom associated with short length-scale backfolds along the chain contour [58, 131, 132, 140, 142]. As stated above, unlike the variance in the extension, the friction diverges as the channel size decreases. Thus for strong confinement, the variance in the mean span dominates the relaxation time.

Since the two curves present in Fig. 7.3 are directly combined to give the relaxation time in Fig. 7.1, we can assess the effect of both the friction and the variance in the extension on the relaxation time. We turn our attention back to Fig. 7.1 to do so. Note that the overall shape of the calculated and experimental relaxation time curves in Fig. 7.1 are similar, including the presence of a maximum in the relaxation time near the Kuhn length. Given our assessment of Fig. 7.3, we postulate that this maximum results from a tradeoff of increasing friction and decreasing variance as the channel size decreases.

7.3.2 Model Assumptions, Limitations and Criticism

Having presented the evidence supporting our modeling approach, we now provide some critical analysis of our assumptions and method. As mentioned, in order to map the detailed WOSM to the one-dimensional dumbbell model, we assumed that small fluctuations about the equilibrium extension were approximately Gaussian. Accordingly, we need information on the probability density function, $\psi(X)$, of the span of the detailed model. Naturally, $\psi(X)$ cannot be exactly Gaussian due to finite extensibility (i.e. X can not exceed L [75]) and hairpin states may cause the distribution to be distinctly non-Gaussian [140, 254]. To avoid any complication due to global hairpins [254], we initialized all of our Monte Carlo runs in extended states. Subsequently, while many backfolded states were realized in weak confinement, no global hairpins were detected for $D \lesssim l_p$. This observation is consistent with the hypothesis that hairpin formation is a rare, slow event in Metropolis Monte Carlo simulations of DNA in tight confinement [140]. However, with this method we are unable to determine if hairpins are in fact absent from the equilibrium ensemble, or if the simulation is incapable of reaching the time scale necessary to observe such configurations. Regardless, it seems likely that experimental measures of the relaxation time also neglect hairpins, since they are easily observed [29, 60, 243, 266, 267].

Figure 7.4(a) shows the resulting probability density functions, $\psi(X)$, obtained from the configurations of the WOSM, where the bulk of the probability density is reasonably well approximated as Gaussian for all channel sizes. Indeed, this approximation is further supported by the normal probability plots shown in Fig. 7.4(b), which show that no significant deviation from Gaussian behavior occurs until ± 2 standard deviations. Since the chain has a finite extension, there must be deviations from Gaussian behavior in the tails of the normal probability plots. Additionally, note that there is poorer sampling in the tails of the distributions, leading to substantial noise in this region.

Perhaps the most uncertain part of our analysis is the estimation of the various parameters l_p, w, a and even L in the presence of the intercalating dye, which is the main reason why we chose to include a fitting parameter in Eq. 7.5. Indeed, it seems that all attempts to quantitatively compare simulations and theory to experimental data for DNA are impacted by the absence of reliable measurements for these parameters for dyed DNA.[237] We have adopted the standard approach here, increasing the contour

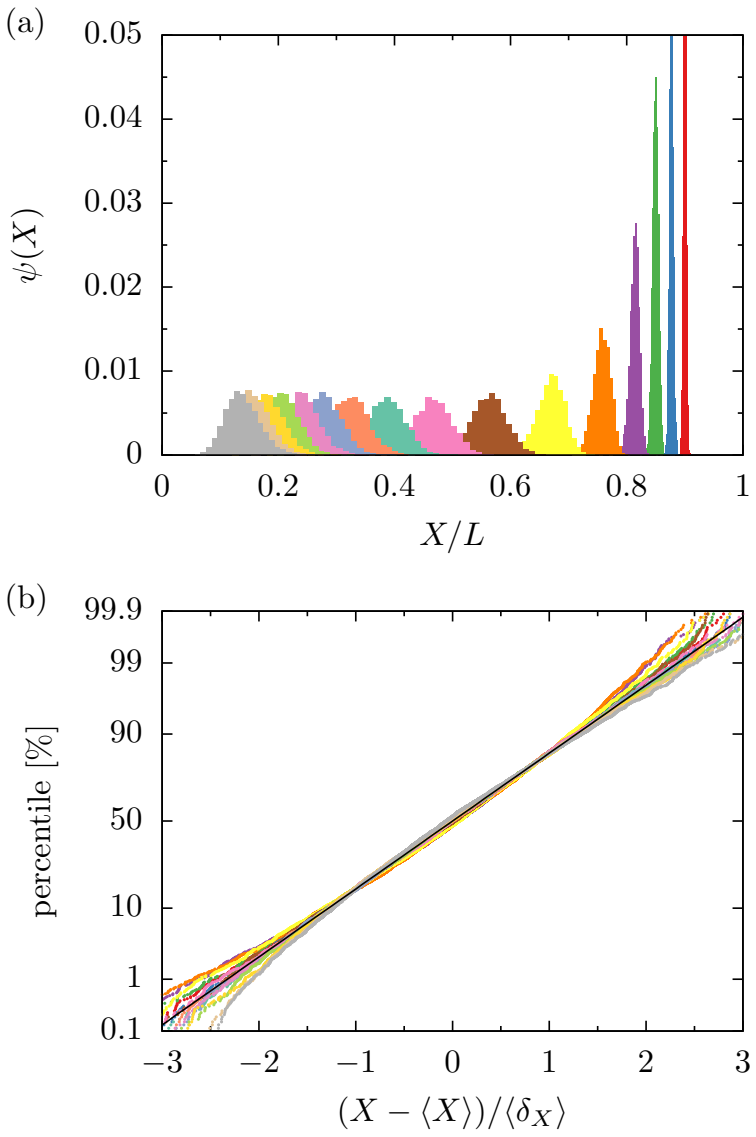

Figure 7.4: (a) Probability density function for the chain extension $\psi(X)$ in the WOSM for different channel sizes as a function of fractional extension. The channel sizes shown are (from right to left in nm): 30, 36, 43, 52, 62, 74, 89, 106, 127, 152, 183, 219, 262, 314, 376, 450. (b) Normal probability plot of the 16 distributions. The solid black line indicates a normal distribution. The color scheme in both panels is the same.

length by 20-30% while assuming the persistence length is unchanged, inspired by the success of these parameters in models for DNA electrophoresis [268].

Given the number of approximations required to reduce an intractable dynamic simulation problem to a feasible Monte Carlo calculation, the quantitative agreement in Fig. 7.1 is satisfying, especially since all of the uncertainty in the parameters and the assumptions required for the methodology seem to be reducible to a single $O(1)$ fitting constant. In particular, while the Kirkwood approximation is known to be accurate for weak confinement [176], there are no experimental data or dynamic simulation data to assess its accuracy in strong confinement. A possible route to improving the quantitative agreement is to modify the WOSM so that the spheres correspond to the hydrodynamic diameter a rather than the excluded width w, thereby improving the accuracy of the hydrodynamic calculations while simultaneously reducing discretization artifacts in the smaller channels. However, the tradeoff is an increased number of beads required to reach a particular contour length L; the requisite hydrodynamics calculations entail improvements in the methodology that are currently in development. We also recognize the need to develop a more sophisticated model for the DNA-wall interactions in the Odijk regime, where the hardcore excluded volume is probably insufficient.

7.4 Conclusion

In the present contribution, we showed how a one-dimensional, overdamped, Rouse dumbbell model, parameterized from detailed Monte Carlo simulations, can reasonably reproduce the relaxation time data observed in experiments [60] using a single $O(1)$ fitting parameter. While there are a number of assumptions underlying our analysis, the final result in Fig. 7.1 suggests that our model is sufficient to capture the extant experimental data. There are two obvious routes to test our approach. One option is to acquire experimental data under the same experimental conditions, since our simulation results make a testable prediction about the shape of the relaxation time curve. However, it seems likely that additional experimental data points would not invalidate our result, especially in light of the experimental uncertainties. A more promising avenue is to acquire experimental data for the relaxation time at different ionic strengths, which will alter simultaneously the persistence length and the effective width (see Chapter 4

and Refs. 66, 269). Inasmuch as we have already used our adjustable parameter in Eq. 7.5, changing ionic strength would provide a stringent test of our approach. Naturally, we would also need to recompute the data in Fig. 7.3 for the new values of l_p and w, but this is a straightforward computational task. If additional experimental data ultimately point out a shortcoming in our result, it is worthwhile to consider what might be the weakest link connecting the detailed Monte Carlo simulations to the relaxation time. Based on Fig. 7.4, it would seem that the assumption of a Gaussian spring for the fluctuations about the equilibrium extension is a possible source of error. While the bulk of the probability distribution is captured by a Gaussian distribution, the tails certainly cannot be Gaussian. Moreover, depending on the amount of weight in the tails, they may make a non-trivial (but still finite) contribution to the variance in the extension. Fortunately, there is already a body of literature on polymer rheology using dumbbell models with more sophisticated spring laws [118]. Moreover, one should be able to construct an appropriate spring by comparing its thermal properties to the histograms in Fig. 7.4. While any such model is certainly more complicated than the Gaussian model we used here, the additional degrees of freedom in a more complicated spring law — along with a finite extensibility — should improve the agreement with experimental data.

Moving beyond the methodology, our results suggest that a dramatic reduction in the fluctuations of the span of the chain as the channel size drops below the Kuhn length of the DNA, not the increased friction, are responsible for a qualitative change in the relaxation time. This phenomenon is not just a scientific curiosity; it is critical to the success of state-of-the-art DNA mapping in nanochannels. In the commercial method [47, 56], the DNA molecules are decorated with sequence-specific probes and injected into an array of nanochannels. The resulting massively parallel array of linearized molecules is imaged in a series of consecutive scans. The extension fluctuations thus set the lower bound for the error in a single snapshot of the extension between two barcodes. The most recently reported nanochannel mapping device [47] uses 45 nm $\times$ 45 nm channels, which are well below the 100 nm Kuhn length of DNA. Our data suggest that these channel sizes suppress the variance in distance between barcodes, thereby permitting a robust identification of the structural variations that are critical to understanding genomic diversity [47].

Chapter 8

The Odijk Regime in Slits

This chapter is based on the publication:
D. R. Tree, W. F. Reinhart, and K. D. Dorfman
Macromolecules (2014), In Review

8.1 Introduction

The subject of confined wormlike polymers has garnered recent attention due to the confluence of fundamental questions in polymer physics and the demand for new genomic analysis tools [51, 270]. One of the simplest modes of confinement, both theoretically and experimentally, involves placing a wormlike chain between two parallel planes. This type of confinement exhibits a rich set of equilibrium behavior due to the interplay between confinement, chain stiffness, excluded volume interactions and chain length [136, 140–143, 146–148]. However, despite the apparently simple geometry, details surrounding strongly confined wormlike polymers in slits remain remarkably controversial [140, 142, 147]. In this article, we attempt to resolve the issue by proposing a comprehensive theory of the Odijk regime in slits, which we subsequently support with Monte Carlo simulations.

The source of disagreement in the literature surrounds a divergence of approach in attempting to resolve the details of the Odijk regime in slits. Using a combination of field theory and simulation, Burkhardt *et. al.* [131, 132, 271] and Chen and Sullivan [240] have examined polymers absent any intra-chain interactions (*i.e.* ideal chains). By different

methods, both groups [131, 132, 240, 271] produce strong, independent evidence for the existence of deflection segments in slits, and Burkhardt *et. al.* [131, 132, 271] provide precise predictions of the relevant numerical prefactors. Alternatively, Doyle *et. al.* [142, 147] and Cifra [140] have studied real chain models, which include excluded volume interactions and aim to directly compare with experiments using DNA. In the latter work, the evidence for the Odijk regime is less clear — so much so that one study suggests that the Odijk regime may not even exist in slit confinement [140].

We reconcile these threads in the literature with a theory for strongly confined polymers in slits based on the two simple postulates that are embodied in the schematic shown in Figure 8.1. We first posit that a sufficiently long, strongly-confined chain may be viewed as a two-dimensional chain with weak swelling in the direction normal to the confinement plane due to deflection segments, as seen in the left panel in Figure 8.1. Such a theory is plausible if the formation of deflection segments is (i) independent of chain properties such as persistence length and excluded volume and (ii) separable from the behavior of the chain along the other dimensions. The fact that the chain is essentially two-dimensional should not be particularly surprising, since it is analogous to the mapping proposed by Daoud and Gennes [59] for blob regimes in weak confinement. Indeed, this analogy is easily visualized if one imagines that the beads of the chain in the left panel of Figure 8.1 are replaced by blobs.

The second postulate concerns the nature of chains with excluded volume and is shown in the right panel of Figure 8.1. When the slight height, H, is much greater than the chain width, w, the probability of intra-chain interactions becomes vanishingly small and the ideal limit is recovered. The opposite happens when $H = w$. Here, the chain cannot avoid itself and becomes equivalent to a purely two-dimensional, self-avoiding wormlike chain in the plane of confinement [208, 272]. The case $H > w$ exhibits some probability of interaction and is thus a more complicated, intermediate state. Using these insights, we propose that a theory that aims to describe slit-confined chains with excluded volume must be consistent with the ideal chain case discussed in the previous paragraph when $H \gg w$ and must also reduce to a real, two-dimensional wormlike chain when $H = w$.

We proceed to incorporate these postulates into a theory of the Odijk regime in slits. First, we extend the ideal chain results of Burkhardt *et. al.* [131, 132, 271] and

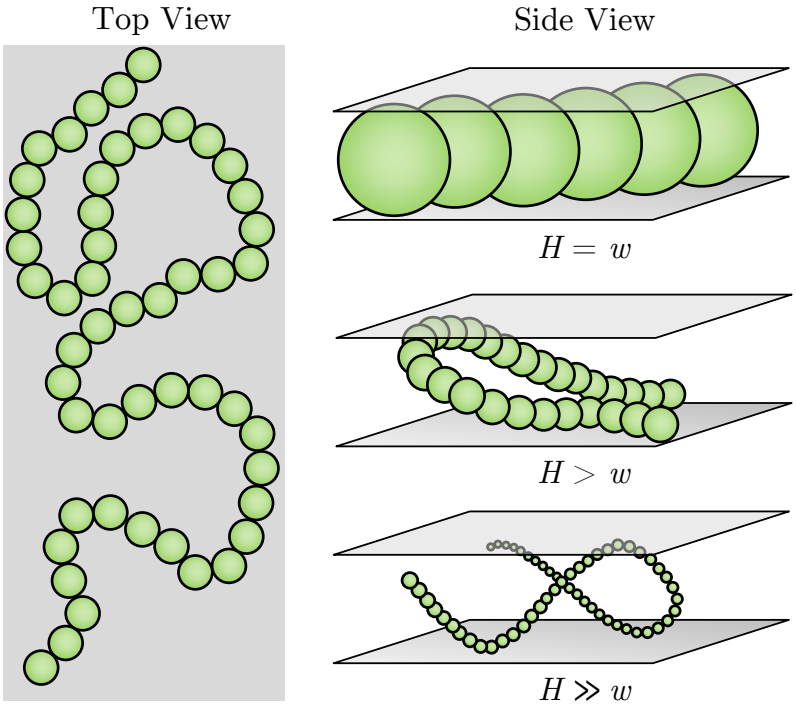

Figure 8.1: Schematic representation of real chains in varying degrees of strong, uniaxial confinement (i.e. $H \ll l_p$). The left presents a top-down view, showing a two-dimensional wormlike chain composed of deflection segments. The right shows a cross-sectional view, which highlights the effect of excluded volume strength. For $H = w$, the polymer becomes a wormlike chain confined to a plane and is a purely, two-dimensional self-avoiding walk. Alternatively, when $H \gg w$ the probability of segment interactions is greatly suppressed and the chain becomes ideal. Note that this schematic does not depict a fixed value of the persistence length, which must be much greater than the slit height in all cases.

Chen and Sullivan [240] into a theory that quantitatively relates a strongly confined chain and a 2D wormlike chain. We then incorporate the effects of excluded volume using a Flory theory originally proposed by Odijk [136], which accounts for both the limits $H \gg w$ and $H = w$ described in Figure 8.1. Having reviewed and expanded on the theory of the Odijk regime in slits, we compare its predictions to numerical results of a discrete wormlike chain model using an off-lattice version of the pruned-enriched Rosenbluth method [171]. The method allows us to explore a large region of the applicable parameter space over which we find excellent agreement with our theory. Finally, we conclude with a brief discussion of some of the implications of our findings.

8.2 Theoretical Results

8.2.1 Ideal Chains

We now consider detailed theoretical results for the equilibrium behavior of an ideal wormlike chain in the Odijk regime in slits. Alongside our own results, we also review prior work on ideal chains, which is necessary for a self-contained discussion of chains with excluded volume in Section 8.2.2. Given the need to review this material and our desire to preserve logical continuity, ideas and equations attributable to other authors are often simply distinguished with a citation. Conversely, an equation with no such citation can be considered an original contribution.

A slit-confined, ideal wormlike chain has three length scales which characterize its behavior: a slit height, H, a contour length L and a persistence length l_p. In the present work, we will focus solely on the region of the parameter space where $H \ll l_p$, which we take as the definition of the Odijk regime. Furthermore, we restrict our scope to proper uniaxial confinement, with an infinite domain extending in the directions tangent to the floor and ceiling of the slit. In practice slits must of course have side walls, but it is anticipated that the walls are unimportant for the equilibrium conformation as long as the polymer size, R, is small compared to the distance between the walls [31, 136].

The Odijk regime is distinguished by the emergence of a new length scale, which manifests itself when $H \ll l_p$ [58]. This length scale, called the deflection length [58]

$$\lambda = H^{2/3} l_p^{1/3} \tag{8.1}$$

describes the contour length of the confined chain between intermittent collisions with the top and bottom of the slit walls (see Figure 8.1). Since the persistence length is large on the scale of confinement, bending of the chain is minimal between such deflections, and a segment of the chain of the size of the deflection length is essentially rod-like.

In addition to the restriction that H be much less than l_p, the contour length of the chain must be sufficiently long to enter the Odijk regime. In fact, given that $H \ll l_p$ we can distinguish between three regimes of behavior based upon the contour length of the chain. When $L \gg \lambda$ we achieve the Odijk regime, where we have sufficient chain length to exhibit a large number of deflection segments. On the other hand, if $L \ll H$ the chain is weakly confined and behaves similar to rod-like particle in the bulk. Due to the separation of length scales between H and λ when l_p is large, there is an intermediate regime which arises when $H \ll L \ll \lambda$. This regime is characterized by orientational ordering of the stiff chain, similar to nematic ordering of rod-like liquid crystals [273].

The transition from bulk behavior to the Odijk regime can be characterized by both a projection of the contour length onto the plane of confinement, $L_\parallel$, as well as a confinement free energy, ΔF_c. We define the projected contour length as the length of the chain contour projected onto the plane parallel to the slit surface

$$L_\parallel = \left\langle \int_0^L \left| \frac{\partial \mathbf{r}_\parallel}{\partial s} \right| ds \right\rangle \tag{8.2}$$

where $\mathbf{r}_\parallel(s) = \mathbf{e}_\perp \times \mathbf{r}(s) \times \mathbf{e}_\perp$ is the projection of the contour of the chain, $\mathbf{r}(s)$, onto the slit surface assuming that $\mathbf{e}_\perp$ is a unit normal to the plane of confinement. The confinement free energy is given by

$$\Delta F_c = k_B T \ln \frac{Z}{Z_0} \tag{8.3}$$

where Z is the configurational partition function, Z_0 is the partition function for an ideal, unconfined wormlike chain, k_B is Boltzmann's constant and T is the absolute temperature.

Focusing on the projected contour length for the moment, we note that it is difficult to obtain an analytical expression from Eq. 8.2, and we are forced to rely on numerical techniques. Fortunately some simple trigonometric arguments are sufficient to obtain approximate expressions for $L_\parallel$ in each regime. When $L \ll H$, the polymer is rod-like,

and we can approximate the projected contour length as

$$L_{\parallel} \approx L\langle \cos\theta \rangle \tag{8.4}$$

where θ is the smaller angle between L and $L_{\parallel}$. Because L is smaller than the slit height, the rod is free to rotate isotropically, giving [273]

$$L_{\parallel} = \frac{\pi L}{4} \tag{8.5}$$

which is what we would expect as bulk-like behavior. In the orientation regime, when $H \ll L \ll \lambda$, the chain is still rod-like but can no longer rotate freely. In this case we can assume that $\theta \approx H/L$ is small and we can expand the right hand side of Eq. 8.4 to give [58]

$$L_{\parallel} \approx L\left[1 - \frac{1}{2}\left(\frac{H}{L}\right)^2\right] \tag{8.6}$$

A more careful evaluation of Eq. 8.2 yields only a change to the prefactor [131, 132, 271]

$$L_{\parallel} = L\left[1 - \alpha_L\left(\frac{H}{L}\right)^2\right] \tag{8.7}$$

with $\alpha_L = 0.09137$. Finally, in the Odijk regime, $L \gg \lambda$ and the chain consists of $N_\lambda = L/\lambda$ deflection segments. Assuming each deflection segment contributes a portion similar to the orientation regime, we substitute $L = \lambda$ into the right-hand side of Eq. 8.7 and multiply by N_λ to obtain [58, 131, 132, 271]

$$L_{\parallel} = L\left[1 - \alpha_L\left(\frac{H}{l_p}\right)^{2/3}\right] \tag{8.8}$$

The confinement free energy also shows distinctive behavior in each of the three regimes. When $L \ll H$, a combination of a loss of translational and rotational entropy gives rise to [239, 274]

$$\Delta F_c = -\ln\left(1 - \frac{L}{2H}\right) \tag{8.9}$$

This free energy is identical to that of a weakly confined polymer in larger slits ($H \gg l_p$) and is exactly half that of a weakly confined polymer in a square channel [239, 275]. When $H \ll L \ll \lambda$, the loss of rotational entropy becomes more severe and [273, 274]

$$\Delta F_c = k_B T \ln\left(\frac{2L}{H}\right) \tag{8.10}$$

When $L \gg \lambda$, dimensional analysis suggests that the confinement free energy ΔF is $\mathrm{O}(k_B T)$ per deflection segment is [58, 273]

$$\Delta F_c \sim k_B T \frac{L}{\lambda} \tag{8.11}$$

Odijk has pointed out that Eq. 8.11 is not equivalent to N_λ independent segments with a free energy given by Eq. 8.10 [58, 273]. Instead, Eq. 8.11 also accounts for the correlation between successive deflection segments due to chain connectivity [273]. Rearranging Eq. 8.11 and introducing a dimensionless confinement free energy per persistence length $\mathcal{F}$, gives [131, 132, 271]

$$\mathcal{F} \equiv \frac{\Delta F_c}{k_B T} \frac{l_p}{L} = A \frac{l_p}{\lambda} \tag{8.12}$$

where $A = 1.1036$ is a geometry-dependent prefactor previously calculated by Burkhardt *et. al.* [131, 132, 271].

It is worth pointing out that since the Odijk regime is asymptotically approached when $H \ll l_p$, the confinement free energy smoothly varies between the strong confinement limit [58] and the weak confinement limit [59, 265] as a function of H/l_p. Accordingly, Chen and Sullivan provide an interpolation formula for the free energy of ideal chains as a function of slit height [240],

$$\mathcal{F} = \frac{\pi^2/3 \left(l_p/H\right)^2}{\left[5.146(l_p/H)^2 + 1.984(l_p/H) + 1\right]^{2/3}} \tag{8.13}$$

valid at all values of H/l_p. Note that Eq. 8.13 reduces to Eq. 8.12 in the limit $H/l_p \ll 1$.

In many experimental systems, $L \gg \lambda$ and the chain is long enough to be safely considered in the Odijk regime. In such systems (*e.g.* fluorescence microscopy experiments involving double-stranded DNA), the polymer size projected onto the confinement plane, $R_\parallel$, is directly observed. For our purposes it is more convenient to deal with the root-mean-square end-to-end distance

$$R \equiv \langle |\mathbf{r}(L) - \mathbf{r}(0)|^2 \rangle^{1/2} \tag{8.14}$$

Using simple trigonometric arguments similar to those above, $R_\parallel$ can be related to R by

$$R_\parallel = R \left[1 - \gamma \left(\frac{H}{R} \right)^2 \right] \tag{8.15}$$

where γ is some unknown constant. Importantly, the ratio $R_{\|}/R$ given by Eq. 8.15 approaches one with increasing chain length regardless of the slit height, unlike $L_{\|}/L$, which remains a function of H/l_p. Therefore, because $R \gg \lambda \gg H$ is a requirement for the Odijk regime, one may safely assume that for most practical purposes $R_{\|}$ and R are interchangeable.

Given its experimental importance, we would certainly like to find an expression for the end-to-end distance as a function of L, l_p and H for an ideal chain, which seems to be absent in the existing literature. As long as $L \gg \lambda$, the formation of deflection segments in strong confinement decouples the behavior of the chain in the directions parallel and perpendicular to the plane of confinement. Thus, in $\mathbf{e}_{\perp}$ the chain configurations are dominated by stochastic collisions of deflection segments with the top and bottom of the slit, irrespective of the horizontal position of the chain. However, since the parallel axes contain no impediments, the chain is free to swell horizontally and perform some type of two-dimensional random walk [136, 142]. As stated in the outset, this assumption is similar to Daoud and de Gennes' blob theory [59], and forms the basis of several Flory theories used to explain the behavior of real chains in the Odijk regime [136, 142]. However, instead of a two-dimensional walk of blobs, in the strong confinement, we have a two-dimensional walk of deflection segments.

Moving to express these ideas more formally, dimensional analysis reveals that the square end-to-end distance is a function of L/l_p and λ/l_p only

$$\frac{R^2}{{l_p}^2} = f\left(\frac{L}{l_p}, \frac{\lambda}{l_p}\right) \tag{8.16}$$

In Eq. 8.16, the dependence of R on slit height is expressed in terms of λ/l_p only, since λ is the only quantity that depends on H. Hypothesizing that the random walk behavior of the chain is decoupled from the slit height leads us to

$$\frac{R^2}{{l_p}^2} = g\left(\frac{L}{l_p}\right) h\left(\frac{\lambda}{l_p}\right) \tag{8.17}$$

Here g accounts for the walk in the plane of confinement and h accounts for the excursions of the chain in $\mathbf{e}_{\perp}$.

We hypothesize that since the chain is wormlike in the bulk, $g(L/l_p)$ also describes a wormlike chain, in contrast to a different type walk such as a random flight. If the chain

is indeed wormlike, g can be described by the Kratky-Porod model for a two-dimensional wormlike chain [208, 272]

$$g\left(\frac{L}{l_p}\right) = \frac{4L}{l_p}\left[1 - \frac{2l_p}{L}\left(1 - e^{-L/(2l_p)}\right)\right] \tag{8.18}$$

As an additional check, one may also evaluate the bond autocorrelation function

$$C(s) = \langle \mathbf{u}(s) \cdot \mathbf{u}(0) \rangle \tag{8.19}$$

where $\mathbf{u} = \partial\mathbf{r}/\partial s$ is a tangent vector to the chain contour. $C(s)$ gives the directional correlation between chain segments as a function of the distance along the backbone of the chain, and for an ideal, 2D wormlike chain is given by [208, 272]

$$C(s) = \exp\left(-\frac{s}{2l_p}\right) \tag{8.20}$$

Notably, the characteristic decay length for a 2D wormlike chain is twice the persistence length l_p of a wormlike chain in bulk solution.

To complete our description of the end-to-end distance, we still need to account for the change in R due to the change in slit height. Because $H \ll l_p$, the dimensionless quantity λ/l_p is small and $h(\lambda/l_p)$ may be expressed as a series expansion

$$h\left(\frac{\lambda}{l_p}\right) = 1 - 2\alpha_R\frac{\lambda}{l_p} + \mathrm{O}\left[\left(\frac{\lambda}{l_p}\right)^2\right] \tag{8.21}$$

where we define α_R to be a positive constant. We expect this approximation to be very good for large l_p, similar to Eq. 8.8 for the projected contour length.

Combining Eq. 8.1, Eq. 8.18 and Eq. 8.21 gives

$$R^2 = 4Ll_p\left[1 - \frac{2l_p}{L}\left(1 - e^{-L/(2l_p)}\right)\right]\left[1 - 2\alpha_R\left(\frac{H}{l_p}\right)^{2/3}\right] \tag{8.22}$$

where we have dropped the second order terms in λ/l_p. In addition to providing a quantitative prediction, Eq. 8.22 implies that we can distinguish between two subregimes based on the projected contour length of the chain. For $L \ll 2l_p$, we find an "Odijk–rod" regime where the chain behaves like a stiff rod in the plane of confinement. Simplifying Eq. 8.22 in the rod limit gives

$$R = L\left[1 - \alpha_R\left(\frac{H}{l_p}\right)^{2/3}\right] \tag{8.23}$$

where the Taylor series expansion $\sqrt{1-\lambda/l_p} \approx 1-\lambda/2l_p$ has been used to approximate the square root function to first order in λ/l_p. Alternatively, when $L \gg 2l_p$, we encounter an "Odijk–Gaussian" regime where the chain obeys a random walk in the plane of confinement. Here Eq. 8.22 simplifies to

$$R = 2(Ll_p)^{1/2}\left[1-\alpha_R\left(\frac{H}{l_p}\right)^{2/3}\right] \tag{8.24}$$

Note that the "rod" and "Gaussian" terminology employed here reference the rod-like and Gaussian-like regimes of dilute-solution ideal wormlike chains, typically associated with the Kratky-Porod model (see Chapters 4 and 5).

8.2.2 Real Chains

Up to this point, we have limited our discussion to ideal chains. We now abandon this assumption and consider a polymer with a hardcore diameter w confined in a slit of height H. Fortunately, the effects of excluded volume interactions are suppressed for very stiff chains and many of the results of Section 8.2.1 are directly applicable. Indeed, if we consider a chain shorter than $2l_p$, intra-chain interactions are negligible, and the equations contained in Section 8.2.1 for the projected contour length, confinement free energy and end-to-end distance remain valid.

For chains longer than $2l_p$, the impact of excluded volume can no longer be neglected. Here, the effects of stiffness on the in-plane walk are diminished, and the conformation becomes much more likely to experience intra-chain interactions. We expect that both the experimentally relevant, end-to-end distance and the confinement free energy of the chain will be sensitive to these interactions. For example, in Section 8.2.1, we demonstrated that for large values of L, the end-to-end distance of an ideal chain scales like $L^{1/2}$ in a subregime we labelled "Odijk-Gaussian." As indicated in Figure 8.1, we expect to recover this regime for real chains when $H \gg w$. However, when H approaches w, we expect the polymer size to scale like $L^{3/4}$, the Flory scaling for a self-avoiding walk in two-dimensions [59, 208].

Two Flory theories have appeared in the literature in an attempt to explain the effect of excluded volume on a long chain in strong, slit-like confinement [136, 142]. Despite their well-known shortcomings in describing free energies [77–79], both Flory theories provide insight into the crossover from Gaussian to self-avoiding behavior.

Dai *et. al.* [142] have proposed a theory based on the idea of mapping a chain to a two-dimensional projection on the plane of confinement. In their model they define a projected contour length equivalent to Eq. 8.8 and an empirically–defined, projected persistence length, $l_{p,\|}$, which varies smoothly between l_p and $2l_p$. Two-dimensional excluded volume interactions are then expressed in terms of an excluded area [142]

$$a_{\text{ev}} = l_{p,\|}^2 + 1.3wl_{p,\|} \tag{8.25}$$

For our purposes, we note that since $w \leq H \ll 2l_p$, the second term in Eq. 8.25 is at least an order of magnitude smaller than the first. This reduces the excluded area to

$$a_{\text{ev}} = l_{p,\|}^2 \tag{8.26}$$

which is the same as the excluded area which Hsu *et. al.* used for a wormlike chain confined to a flat plane [208].

Using Eq. 8.26, one may write a Flory free energy describing a virtual two-dimensional chain [142, 208]

$$\frac{F}{k_BT} \sim \frac{R_\|^2}{L_\| l_{p,\|}} + N_{l_{p,\|}}^2 \frac{l_{p,\|}^2}{R_\|^2} \tag{8.27}$$

where $N_{l_{p,\|}} = L_\|/l_{p,\|}$ is the number of persistence lengths of the virtual chain. By minimizing Eq. 8.27 with respect to $R_\|$ one finds a polymer size

$$\frac{R_\|}{l_{p,\|}} \sim \left(\frac{L_\|}{l_{p,\|}}\right)^{3/4} \tag{8.28}$$

which gives the needed 3/4 scaling for a self-avoiding walk. While the latter Flory theory is appealing, Hsu *et. al.* have elegantly shown that the theory embodied in Eq. 8.27 lacks a crossover to $L_\|^{1/2}$ for any chain length or excluded volume strength [208]. As such, we are unable to recover the ideal chain limit for $H \gg w$ using such a model.

An alternative Flory theory by Odijk postulates an excluded volume based on isotropic interactions between $N_\lambda = L/\lambda$ deflection segments [136],

$$v_{\text{ev}} = \lambda^2 w \tag{8.29}$$

As we will see, the three-dimensional nature of the excluded volume expressed in Eq. 8.29 gives a crossover with an explicit dependence on w/H.

Following Odijk's approach, we write [136]

$$\frac{F}{k_B T} \sim \frac{L}{\lambda} + \frac{R^2}{2Ll_p} + \frac{L^2 w}{R^2 H} \tag{8.30}$$

where Eq. 8.30 includes an additional term for the confinement of deflection segments (see Section 8.2.1). Note that the L/λ term dominates the others, but falls out when differentiating with respect to R [276]. It also is convenient for a future comparison to include a factor of 2 alongside l_p in Eq. 8.30, since $2l_p$ is the correlation length for a 2D wormlike chain. Differentiating and solving for the in-plane end-to-end distance yields

$$\frac{R}{l_p} \sim \left(\frac{L}{2l_p}\right)^{1/2} \left(\frac{Lw}{2l_p H}\right)^{1/4} \tag{8.31}$$

This result immediately shows the desired dependence on w/H, and we follow Odijk in defining an excluded volume parameter [136]

$$z \equiv \left(\frac{Lw}{2l_p H}\right)^{1/2} \tag{8.32}$$

Thus when $z \gg 1$ the polymer size of the scales like $L^{3/4}$ and when $z \ll 1$, the polymer shows ideal scaling

$$\frac{R}{l_p} \sim \left(\frac{L}{2l_p}\right)^{1/2} \tag{8.33}$$

Comparing the result for ideal chains in Odijk's theory in Eq. 8.33 with Eq. 8.24 from Section 8.2.1 is instructive, since they must agree when $w \ll H$. Here we find that Odijk's theory agrees in part with Eq. 8.24. Recall that Eq. 8.24 was a result of a decomposition into two functions $R = g(L/l_p)h(\lambda/l_p)$, where g accounted for the random walk and h accounted for the fluctuations of the deflection segments in the axis normal to the plane of confinement. Aside from an unimportant difference in prefactor, Eq. 8.33 agrees with the function $g(L/l_p)$, but neglects the term due to $h(\lambda/l_p)$. In other words, Odijk's Flory theory correctly accounts for the physics of the two-dimensional walk only and neglects the (admittedly small) effect due to the change in deflection length when the slit height changes. We postulate that the same decomposition can be done for chains with excluded volume

$$\frac{R^2}{{l_p}^2} = \tilde{g}\left(\frac{L}{l_p}, \frac{w}{H}\right) \tilde{h}\left(\frac{\lambda}{l_p}\right) \tag{8.34}$$

Using Eq. 8.31 to obtain $\tilde{g}$ and our analysis from Section 8.2.1 to obtain $\tilde{h}$ gives

$$R = c(Ll_p)^{1/2}\left(\frac{Lw}{2l_pH}\right)^{1/4}\left[1-\alpha_R\left(\frac{H}{l_p}\right)^{2/3}\right] \tag{8.35}$$

for $z \gg 1$, where c is some O(1) prefactor.

In addition to the effect on R, excluded volume interactions should also impact the free energy of confinement for strongly confined chain. Even with the aforementioned shortcomings in describing free energies [77–79], Odijk's Flory theory expressed in Eq. 8.30 provides considerable insight. Substituting Eq. 8.31 into Eq. 8.30 gives

$$F \sim \frac{L}{\lambda} + \left(\frac{Lw}{2l_pH}\right)^{1/2} \tag{8.36}$$

which shows two terms. The first term is due to the confinement of an ideal chain derived in Section 8.2.1. The second term is due to binary collisions between deflection segments and is equivalent to z. It is the second term that is problematic, due to the neglect of chain connectivity in the mean field approximation inherent in the Flory theory [78, 79]. Instead, directly analogous to a self-avoiding walk, this term becomes linear in the contour length with some unknown functional dependence on the excluded volume strength [78, 79]. Defining an excess free energy as the free energy difference between the confined chain with excluded volume and the confined ideal chain, one can formally express this as

$$\Delta F_{\text{ex}} \equiv F - F_{\text{id}} = \frac{L}{l_p}\mu_{\text{ex}}\left(\frac{w}{H}\right) \tag{8.37}$$

where $\mu_{\text{ex}}(w/H)$ is the excess chemical potential.

In summary, due to the influence of excluded volume interactions, we can now identify a third subregime within the Odijk regime, in addition to the regimes discussed in Section 8.2.1. We label this regime an "Odijk-Flory" regime, in analogy to the "Odijk-Gaussian" regime for ideal chains. Here, chains longer than Hl_p/w show swollen coil behavior with a size given by Eq. 8.35 and an excess free energy given by Eq. 8.37. For the reader's convenience all of the regimes and subregimes discussed in Section 8.2 are summarized in Table 8.1.

Table 8.1: Regimes of strongly confined chains in slits

	bulk	orientation	Odijk		
			Rod	Gaussian	Flory
	$L \ll H$	$H \ll L \ll \lambda$	$\lambda \ll L \ll 2l_p$	$2l_p \ll L \ll l_pH/w$	$l_pH/w \ll L$
$L_\parallel$	$\frac{\pi L}{4}$	Eq. 8.7		Eq. 8.8	
R	L	L	Eq. 8.23	Eq. 8.24	Eq. 8.35
$\beta\Delta F_c$	Eq. 8.9	Eq. 8.10		$\frac{L}{\lambda}$	$\frac{L}{\lambda} + \frac{L}{l_p}\mu_{\text{ex}}$

8.3 Simulation Results

8.3.1 Model and Methods

Having laid out a theory for the Odijk regime in slits, we examine its validity by numerical simulation. To model semiflexible chains, we use a discrete wormlike chain model (DWLC), which is a straightforward discretization of the continuous wormlike chain model used in Section 8.2 and is described at length in Chapters 2–5 and Refs. 71, 189, 190. The main advantage of such a fine-grained model is the ability to resolve sub-persistence length features, which are critical to the study of polymers in the Odijk regime, since deflection segments may be substantially smaller than the persistence length [58]. In our implementation, the polymer is described as a series of beads connected by N bonds of fixed length a giving a contour length $L = aN$. Let the vector $\mathbf{r}_i = (x_i, y_i, z_i)$, $i \in [1, N+1]$ describe the positions of the beads along the contour and let $\mathbf{u}_i = (\mathbf{r}_{i+1} - \mathbf{r}_i)/a$ be the unit vector of the i^{th} bond. In the DWLC model a bending potential is enforced between consecutive bonds

$$U_{\text{bend}} = k_B T \kappa \sum_{i=1}^{N-1} (1 - \mathbf{u}_{i+1} \cdot \mathbf{u}_i) \tag{8.38}$$

with a bending elasticity κ. The persistence length may be expressed in terms of κ by considering the equilibrium probability density function for the bond angle of an ideal

wormlike chain derived in Chapter 4

$$\frac{l_p}{a} = \frac{1}{2}\frac{\kappa - 1 + \kappa \coth \kappa}{\kappa + 1 - \kappa \coth \kappa} \tag{8.39}$$

Note that we take the relation between the bending constant and the persistence length embodied in Eq. 8.39 as the definition of the persistence length, rather than the characteristic length scale of decay of the correlation between successive bonds. As has been discussed by Hsu *et. al.* [277], these definitions are only equivalent when the chain has no excluded volume, and we find that Eq. 8.39 conveniently separates the effects of stiffness due to bending elasticity and stiffness due to proximal intra-chain interactions.

We consider both ideal and real chains, where real chains are endowed with excluded volume interactions between chain segments. This is accomplished by considering spherical beads located at bond joints with hard core repulsion potential described by

$$U_{\text{ev}}(\mathbf{r}_{ij}) = \sum_{i=1}^{N}\sum_{j=1}^{i-1}\begin{cases} \infty, & |\mathbf{r}_{ij}| \leq w \\ 0, & |\mathbf{r}_{ij}| > w \end{cases} \tag{8.40}$$

where w is the bead radius and $|\mathbf{r}_{ij}|$ is the distance between the centers of beads i and j. While it appears from Eq. 8.40 that w is a freely varying parameter, unphysical chain-crossing occurs when w is less than the bond length a. Accordingly, we choose to set $w = a$, resulting in a chain composed of osculating hard spheres. Notably, the choice of a hard core potential also increases the speed of the algorithm.

Both ideal and real chains must also interact with the slit walls. Similar to the excluded volume interaction, chains interact with confinement via a hard potential

$$U_{\text{wall}}(\mathbf{r}) = \sum_{i=1}^{N}\begin{cases} \infty, & |\mathbf{r}_i \cdot \mathbf{e}_\perp| \geq H/2 \\ 0, & \text{otherwise} \end{cases} \tag{8.41}$$

where $\mathbf{r}_i$ is the vector from the origin to the center of bead i, and H is the slit height. Note that Eq. 8.41 confines the centerline of the chain and that this potential corresponds to a physically larger channel for a polymer with excluded volume interactions. For our hard-bead model, this means that the "true" slit height is $H + w$, if we assume that the bead–wall interaction is the same as the bead–bead interaction [71, 131, 132, 271]. A hard potential is of course an approximation, and in a real experimental system both the

interaction potential and the notion of an effective slit height become more complicated. We expect this approximation to break down as the length scales w and H converge, and therefore recommend caution in interpreting our results when $w \approx H$. Notice further that the notation for the slit height used here is different than in our previous work where the true slit height was denoted by H and centerline confinement was given as an effective slit height, H_{eff} [71, 275] (see also Chapters 5 and 7).

We calculate equilibrium properties of the DWLC model using an off-lattice implementation of the pruned-enriched Rosenbluth method (PERM) [171] as described in Chapters 2–5. PERM is a chain growth Monte Carlo algorithm with many advantages over conventional Markov-chain algorithms for very long chains in confinement, since the relaxation time of chains in the latter class of algorithms scale as $\mathrm{O}(N^2)$. Additionally, PERM natively obtains properties as a function of chain length, which is especially useful for the problem at hand.

In the present manuscript, self-avoiding chains are grown up to $5 \times 10^4 + 1$ beads on a DELL Linux cluster using a master/slave parallel algorithm as in Chapters 4 and 5. Static properties were calculated using $\approx 10^5$ tours, which is sufficient to ensure that the standard error of most properties are smaller than the plotted symbol size, unless otherwise noted. In order to streamline future discussion, Table 8.2 outlines the specific data sets referenced in Sections 8.3.2 and 8.3.3.

We are specifically interested in calculating polymer properties relevant for the theory of strongly confined chains presented in Section 8.2. Most of the properties of interest are calculated in a straightforward manner by discretizing their previously given definitions. However, there is some art to efficiently calculating static properties in PERM and further details can be found in Chapters 4 and 5 as well as Appendix A and Appendix B.

For clarity's sake, a brief discussion of the free energy calculations in PERM is also warranted. In PERM, the free energy

$$\Delta F(n) = k_B T \ln \frac{\langle W(n) \rangle}{\langle W_0(n) \rangle} \tag{8.42}$$

is obtained from a ratio of the Rosenbluth weight of the n^{th} bead of the chain $W(n)$ and the Rosenbluth weight of an ideal, wormlike chain in the bulk $W_0(n)$ [171, 275]. In other words, the reference state for the free energy is an ideal, wormlike chain. To

Table 8.2: PERM data sets

	Confined Ideal Chains			
	L/a	l_p/a	H/a	# Tours $\times 10^5$
set 1	[1, 50001] [a]	5000	(5, 500) [b]	10.5
set 2	[1, 50001]	(50, 5000)	5	25.5
set 3 [c]	[1, 50001]	$(10^{1.5}, 10^{3.5})$	(1, 100)	5.5
set 4 [d]	[1, 50001]	{5, 50, 500, 5000}	(50, 500)	0.5
	Confined EV Chains			
	L/w	l_p/w	H/w	# Tours $\times 10^5$
set 5	[1, 50001]	(50, 5000)	5	5.5
set 6 [c]	[1, 50001]	$(10^{1.5}, 10^{3.5})$	(1, 100)	11
	Unconfined EV Chains			
	L/w	l_p/w		# Tours $\times 10^5$
set 7	[1, 50001]	$(10^{1.5}, 10^{3.5})$		5.5

[a] Linearly spaced ranges are denoted with square braces; [b] Logarithmically spaced ranges are denoted with round braces; [c] Data vary such that H/l_p is a constant value at $10^{-1.5}$; [d] Runs were performed for the given range of H/l_p at each of these values of l_p/a.

obtain the confinement free energy of an ideal chain, this reference state is sufficient, and ΔF_c can be obtained from Eq. 8.42 in a single simulation.

The introduction of excluded volume interactions complicates the picture however, since the confinement free energy for a real chain has a real chain reference state. In this case, the free energy of confinement requires a simulation of a chain with excluded volume in the bulk in addition to the confined chain calculation. Therefore, we first perform a simulation of a confined chain with excluded volume and use Eq. 8.42 to obtain $\Delta\tilde{F}$. Then, we evaluate Eq. 8.42 in a simulation of a real chain in the bulk to obtain $\Delta\tilde{F}_0$. The confinement free energy of a chain with excluded volume is then obtained as a difference between these free energies

$$\Delta\tilde{F}_c = \Delta\tilde{F} - \Delta\tilde{F}_0 \tag{8.43}$$

To obtain the excess free energy of a confined chain a similar subtraction procedure is performed to change the reference state to a confined, ideal chain. Here, the excess free energy is the difference between the result of a simulation of a confined chain with excluded volume and a confined, ideal chain

$$\Delta F_{\text{ex}} = \Delta\tilde{F} - \Delta F_c \tag{8.44}$$

8.3.2 Ideal Chains

We are now prepared to compare results from PERM simulations to the predictions contained in Section 8.2 for the Odijk regime in slits. As was the case in Section 8.2.1, in this section we neglect any effects due to excluded volume and examine ideal chains.

We begin by attempting to verify the existence of the weak confinement and orientation regimes for $L \ll H$ and $H \ll L \ll \lambda$ respectively. To this end, Figure 8.2(a) shows the projected contour length for various values of H/l_p as a function of chain length. The agreement with both the functional form and prefactors of the theory in Section 8.2.1 is excellent. As expected the projected contour length follows Eq. 8.5 for $L \ll H$, transitions near $L \approx H$ and then follows Eq. 8.7 for $L \gg H$. The data fall off of Eq. 8.7 at even larger values of L, presumably as the chain enters the Odijk regime. This presumption is supported by the fact that chains with smaller H/l_p values follow Eq. 8.7 for longer chain lengths, suggesting that the width of the orientation regime is governed by the deflection segment length.

Figure 8.2(b) shows a similar story for the confinement free energy versus contour length. For chains much smaller than H the data show excellent agreement with Eq. 8.9, again without any fitting parameters. The data transition at $L \approx H$ and follow Eq. 8.10 for chains much greater than H. As is the case for the projected contour length, the width of the orientation regime depends on the ratio of H/l_p. This observation again supports the fact that the upper bound is determined by the deflection segment length, as discussed in Section 8.2.1.

To examine the transition to the Odijk regime more closely, we focus on Figure 8.3(a), which rescales the data shown in Figure 8.2(a) for the projected contour length as a function of chain length. The data in the figure are again consistent with the ideal chain theory presented in Section 8.2.1. Just as in Figure 8.2(a), the portion of the curve for $L \ll \lambda$ agrees quantitatively with the theory for the orientation regime in Eq. 8.7. Around $L \approx \lambda$ the data transition towards Odijk scaling, and for $L \gtrsim 10\lambda$ it appears to have reached the "long-chain limit." In this limit, when $L \gg \lambda$, the data converge to a value of $\alpha_L = 0.09137$, consistent with prior calculations by Burkhardt *et. al.* [131, 132, 271].

In addition to the data as a function of chain length, Figure 8.3(b) extracts the longest chain lengths in panel (a) and replots them as a function of H/l_p. The agreement between these data and Eq. 8.8 explains why the data in Figure 8.3(a) collapse for different channel sizes. Furthermore, it appears that Eq. 8.8 is a good description of the data even for the largest channel sizes in the data set ($H/l_p \approx 0.1$). Therefore, at least for the projected contour length, we can consider the chain to be in the Odijk regime even for a rather aggressive value of $H \approx 0.1 l_p$.

As is the case with the projected contour length in Figure 8.3, Figure 8.4(a) shows data for the confinement free energy as a function of contour length. The results for the free energy are congruent with both the results for the projected contour length and the theory in Section 8.2.1. Unlike Figure 8.2(b), the data for the free energy of confinement in Figure 8.4(a) are normalized by the number of deflection segments. This causes a peak in the data at $L \approx H$, where the entropic penalty of adding an additional monomer is greatest due to a loss of rotational degrees of freedom. While the peak in Figure 8.4(a) is at $L \approx H$, it appears to move as H/l_p changes; this effect is caused by scaling the x-axis with respect to λ. As expected, the free energy becomes extensive

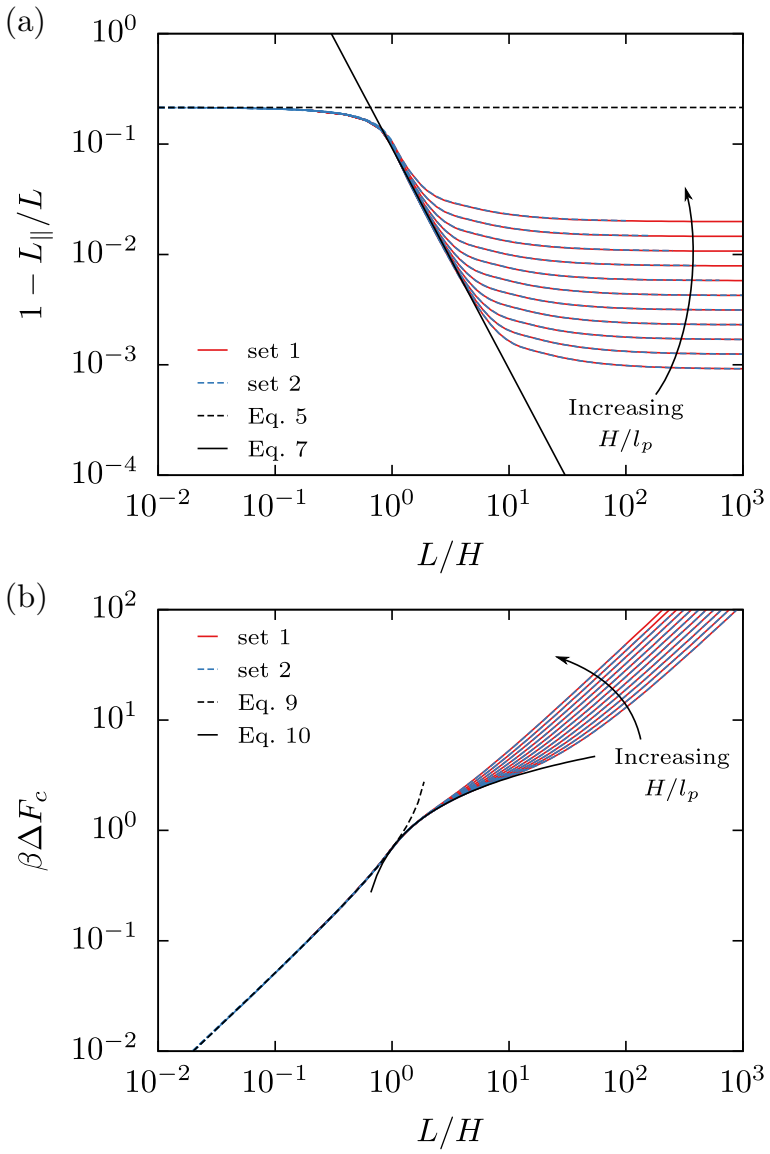

Figure 8.2: PERM data for (a) the projected contour length of slit-confined, ideal chains and (b) the confinement free energy of slit-confined, ideal chains as a function of chain length. In both panels, simulations are run with either a fixed slit height (solid lines) or a fixed persistence length (dashed lines) which corresponds to set 1 and set 2 in Table 8.2.

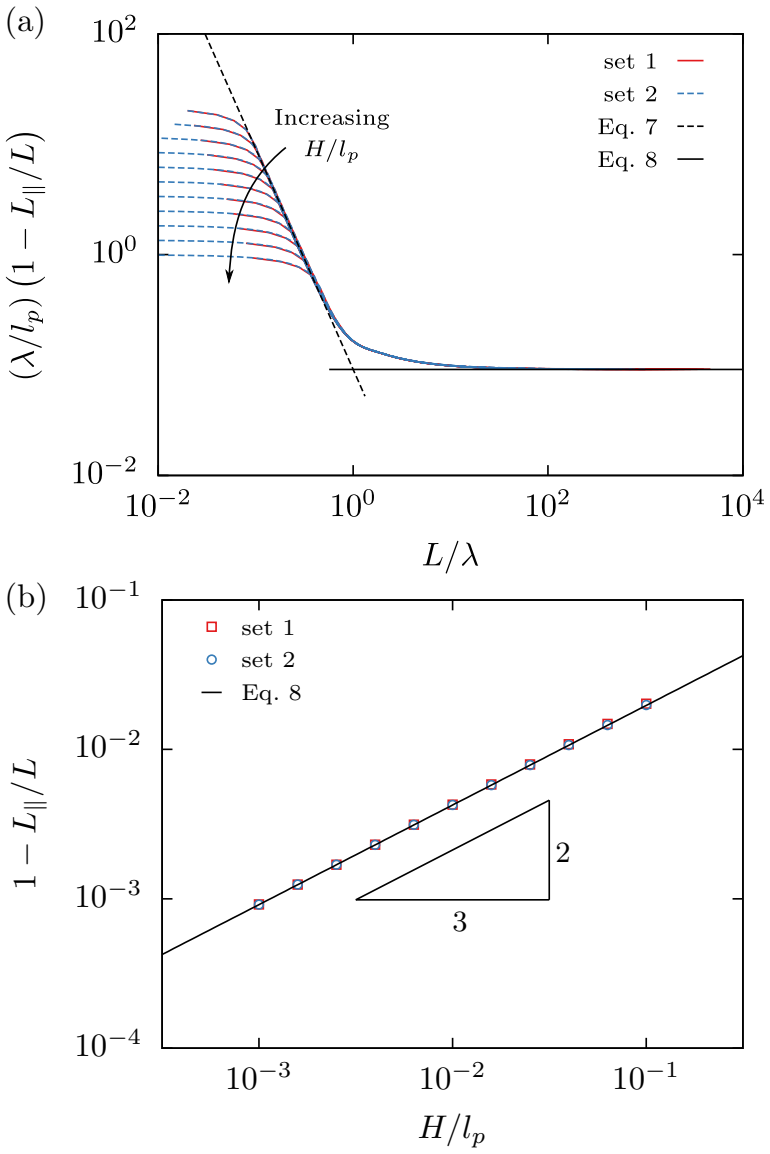

Figure 8.3: PERM data for the projected contour length of a slit-confined chain versus (a) chain length and (b) normalized slit height (for the largest L only). Simulation data sets include data at a constant slit height (squares and solid lines) and a constant chain persistence length (circles and dashed lines). These data correspond to sets 1 and 2 in Table 8.2 respectively. The dashed line in the main portion of panel (a) is given by Eq. 8.7 and the solid curve in (a) and (b) is given by Eq. 8.8. Because of the rescaled y-axis in panel (a), Eq. 8.8 reduces to a constant value of 0.09137.

and proportional to k_BT in the long chain limit. It is a satisfying check of our intuition that the free energy is almost exactly k_BT per deflection segment in this limit. As one might also suspect, the free energy approaches the long chain limit asymptotically, and for most practical purposes, the deviation from the asymptote becomes small when $L \gtrsim 10\lambda$.

For very long chains in Figure 8.4(a), there is some observable spread in the data that appears to disagree with the prefactor of 1.1036 obtained from Burkhardt *et. al.* [131, 132, 271]. This apparent disagreement is due to the finite values of H/l_p used in the PERM simulations. The inset to Figure 8.4(a) shows the largest chain lengths in the main panel, replotted versus H/l_p and compared to the prefactor from Burkhardt *et. al.*. The inset shows that for values of $H/l_p \lesssim 10^{-2}$, the PERM data agree very well with a value of 1.1036, but slit heights larger than this deviate below the curve. Thus in contrast to the projected contour length, the free energy is more sensitive to the slit height and one may need to consider smaller ratios of H/l_p to observe Odijk scaling behavior. This can be understood by recognizing that the prefactor for the free energy is an order of magnitude larger than the corresponding prefactor for the projected contour length. Thus terms of order $(\lambda/l_p)^2$ may also make a substantial contribution to the free energy at the values of H/l_p in the simulation data set.

Figure 8.4(b) examines the dependence of the free energy on H/l_p even further. Here, simulation data for the confinement free energy per persistence length for chains of length $L/a = 50001$ are shown versus a much larger range of H/l_p than in panel (a). When the data is plotted this way, one can compare to the interpolation formula of Chen and Sullivan [240] (Eq. 8.13) in addition to Eq. 8.12. For larger slit heights, the PERM data show a smooth transition from Odijk scaling when $H \ll l_p$ to ideal scaling when $H \gg l_p$, similar to the results of Chen and Sullivan using polymer field theory [240]. Across the entire range of confinement strengths the PERM data agree very well with Eq. 8.13. When the free energy is normalized with the persistence length, $H/l_p \approx 0.1$ appears to denote the onset of Odijk scaling, in contrast to the results in the inset of panel (a). This discrepancy is due to the difference in scaling of the y-axis, which in panel (a) divides by the small term λ/l_p.

We now turn our attention to the theory in Section 8.2.1 for the projected size of the confined chain. Recall that in Eq. 8.15, we used a simple trigonometric argument to

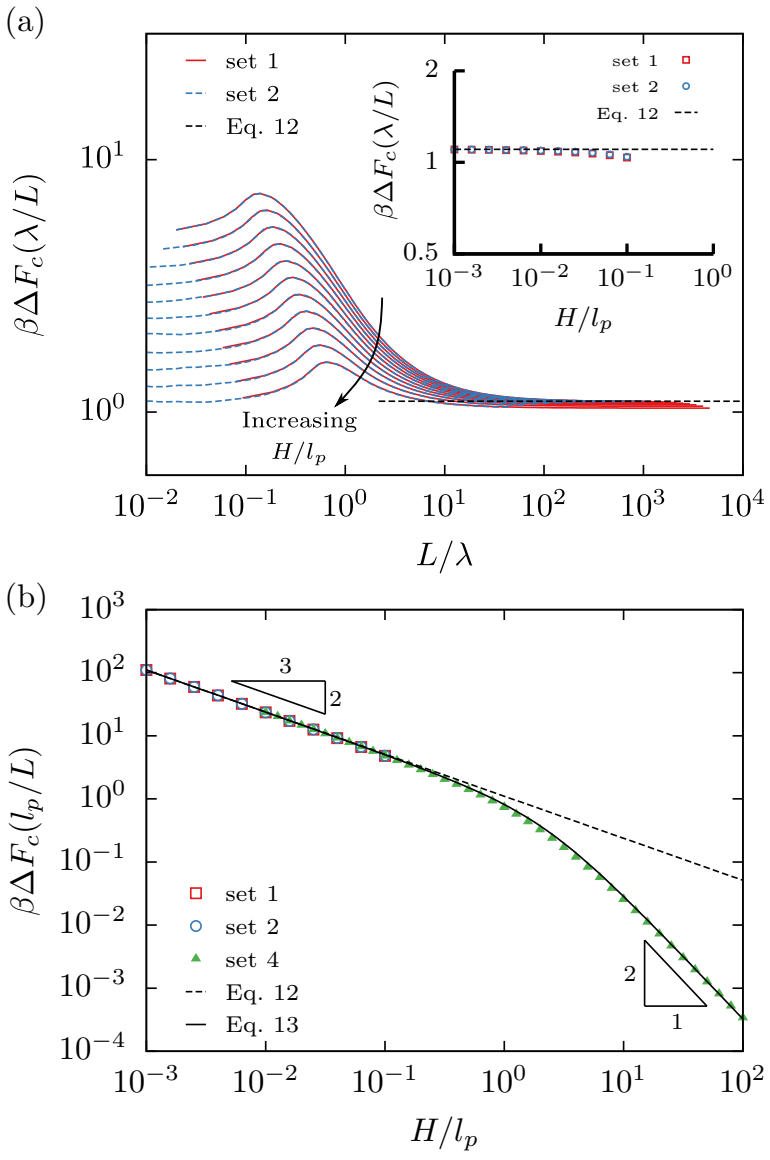

Figure 8.4: PERM data for confinement free energy of a slit-confined chain (a) normalized by the number of deflection segments and (b) normalized by the number of persistence lengths. The data in the main panel (a) are shown versus the number of deflection lengths and versus H/l_p in the inset. The data in panel (b) are shown versus H/l_p. Simulation data sets include data at a constant slit height (squares and solid lines), constant chain persistence length (circles and dashed lines), and variable slit height and persistence length (filled triangles). These data correspond to sets 1, 2 and 4 in Table 8.2 respectively. Note that the dashed line in panel (a) corresponds to Eq. 8.12, which reduces to a constant value of 1.1036 [131, 132, 271].

claim that the projected size, $R_{\|}$, was equivalent to the conventional end-to-end distance for large R. Figure 8.5 shows PERM data for the relative error in this approximation versus R/H. For small values of R/H the relative error remains $\mathrm{O}(10^{-1})$ but falls precipitously for $R \gg H$. The slope of -2 evident on the right hand side of the figure is consistent with the prediction from Eq. 8.15. From the figure, it is clear that $R \approx R_{\|}$ is excellent for values of $R \gtrsim 10H$.

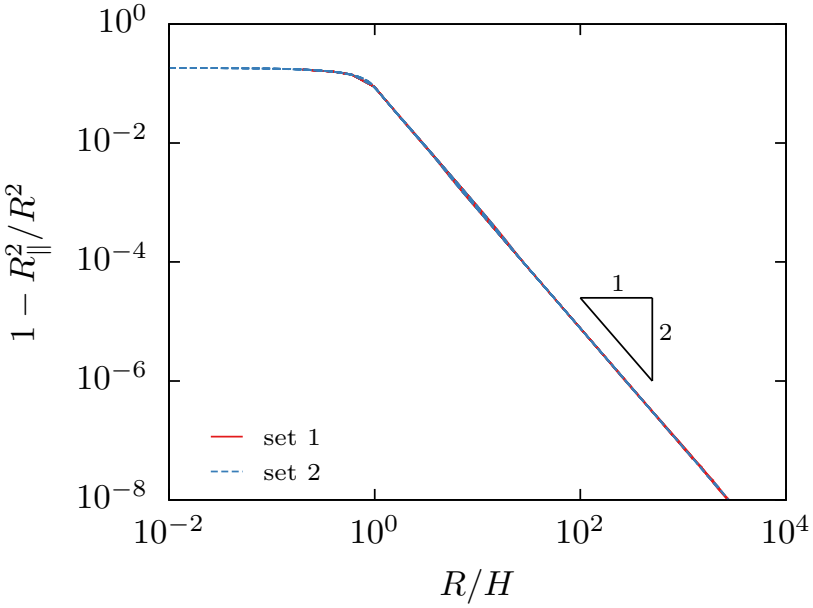

Figure 8.5: Simulation data for the relative error between $R_{\|}$ and R as a function of the rms end-to-end distance of the chain.

Having established that R is a reasonable metric, we would like to validate the key assumptions of our theory. Recall that in Section 8.2.1 we postulated that the function describing R could be decomposed into some function $g(L/l_p)$ describing a random walk and another function $h(\lambda/l_p)$ to account for the deflection segments. Focusing on $g(L/l_p)$ for the moment, we postulated that the confined chain could be approximated as a two-dimensional walk of persistent segments, *i.e.* a 2D wormlike chain.

To test this hypothesis, Figure 8.6 shows simulation results for the bond autocorrelation function $C(s)$ for simulations at various values of H/l_p. In order to collapse the data to a single curve, the autocorrelation function in Figure 8.6 has been rescaled by the power l_p/λ. The collapse is very good, and as expected from Eq. 8.20, the

bond correlations for $s \gg \lambda$ decay exponentially with a characteristic length of $2l_p$. This is indicative of a two-dimensional wormlike chain and provides confirmation of our hypothesis.

However, as was the case with the confinement free energy and the projected contour length, there is also some nuance to the behavior of $C(s)$ as a function of the distance along the contour. In Figure 8.6, we observe that for $s \ll \lambda$ the correlations decay with a characteristic length of l_p, exactly half of the value observed for large s. Of course, the relationship $C(s) = \exp(-s/l_p)$ shown by the dashed line in the figure is exactly what one would expect for an unconfined, three-dimensional wormlike chain. Thus, the chain is experiencing correlations indicative of bulk, 3D behavior on length scales much less than λ and confined, 2D behavior on length scales much greater than λ. In addition to the interest to the problem at hand, this result provides a satisfying example of an oft-made assumption that properties below the scale of confinement can be treated as effectively bulk-like.

Additionally, the results encapsulated in Figure 8.6 are somewhat at odds with previous work, which discuss bond correlations in terms of a projected persistence length, $l_{p,\parallel}$, that depends on H/l_p [142]. However, there are at least two good explanations as to why our data disagree. First, up to this point, we have not considered the effect of excluded volume as was done in prior work. Indeed, Hsu *et. al.* have shown that simple relationships between various definitions of the persistence length break down upon inclusion of excluded volume [277]. Secondly, and more importantly, the authors of the previous study consider a much wider range of values of H/l_p, whereas we limit ourselves to the Odijk regime, where $H \ll l_p$. If we also restrict the equations provided in Ref. 142 to the Odijk regime, we find that Eq. 8.20 and the results in Figure 8.6 are both in good agreement [142].

Having confirmed that the random walk is indeed wormlike, we proceed to test the predictions of Eq. 8.18 for $g(L/l_p)$ versus the values of R obtained form our PERM simulations. Accordingly, Figure 8.7(a) shows PERM simulation data compared to Eq. 8.18 for a range of values of H/l_p. Note that Eq. 8.18 is simply the result for the end-to-end distance of a 2D wormlike chain [208, 272] and does not include any dependence on the slit height. The data agree very well with Eq. 8.18, and the curves are practically indistinguishable except for some notable sampling error that accompanies simulations

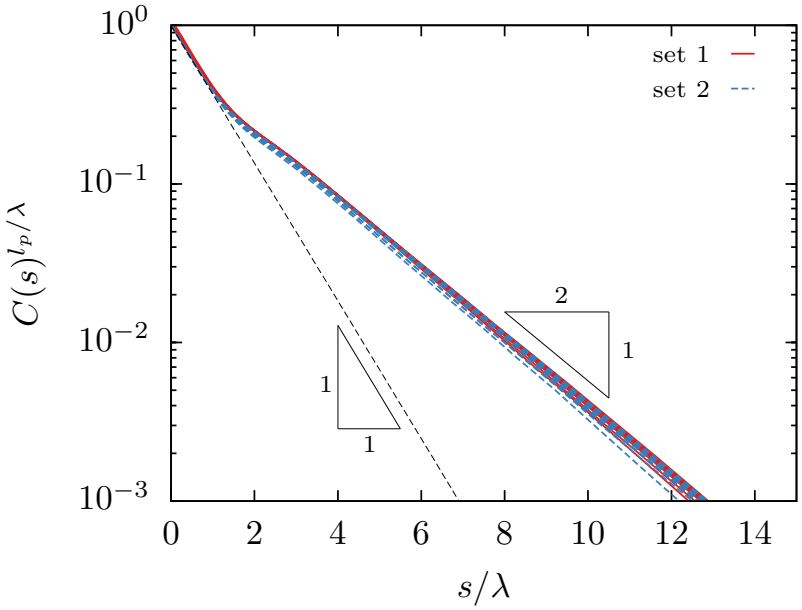

Figure 8.6: Simulation data for the bond autocorrelation function $C(s)$ versus contour distance s for various values of H/l_p. The data presented in this figure correspond to sets 1 and 2 in Table 8.2. The autocorrelation function has been rescaled by a power of l_p/λ in order to collapse all of the data to a single curve. The dashed line depicts the curve $C(s) = \exp(-s/l_p)$.

of extremely large chains in PERM. In fact, at first glance it appears that the simulation data are in quantitative agreement with the prediction even without accounting for the dependence on H/l_p.

To investigate this further, the inset in Figure 8.7(a) shows the relative error between R^2 and g by plotting

$$1 - \frac{R^2}{l_p^2}\frac{1}{g} = 1 - h\left(\frac{\lambda}{l_p}\right) \tag{8.45}$$

as a function of the number of deflection segments for different values of H/l_p, where we have used Eq. 8.17 as our definition for h. As anticipated, there is a small, but systematic deviation in the data as a function of H/l_p. To explain this deviation, we anticipated in Section 8.2.1 that h could be expressed as the series expansion given by Eq. 8.21. Using Eq. 8.21 as a guide, we rescale the relative error by by λ/l_p and obtain Figure 8.7(b). The collapse of the data is excellent (again with some notable sampling error for long chains), showing that the first-order terms are sufficient to capture the dependence of R^2 on H/l_p. Additionally, a fit to the data in the range $L/\lambda \in [1, 10]$ yields a value of $\alpha_R = 0.1039(5)$, which completes the expression for the square end-to-end distance in Eq. 8.22.

Since Eq. 8.22 appears to quantitatively describe the end-to-end distance, it should estimate the in-plane size from a polymer's free solution molecular weight, persistence length and the slit height. This should be an excellent approximation for long chains, since we believe the difference to be minimal between $R_\parallel$ and R when $R \gg H$. In addition, since the value of α_R is $\mathrm{O}(10^{-1})$ and $\lambda/l_p \ll 1$, Eq. 8.18 should also be an excellent approximation for $R_\parallel$, as is evident in Figure 8.7(a). Other useful approximations are contained in Equations 8.23 and 8.24 which contain assumptions about the contour length as described in Section 8.2.1. These equations are also displayed in Figure 8.7(a).

Finally, we close this section with a comment about an alternative theory, which we label the "projected 2D chain theory" [142]. In this theory, the statistics of the three-dimensional chain are interpreted in terms of a mapping to a two-dimensional chain. The parameters of the mapping are $L_\parallel$ and $l_{p,\parallel}$, the contour length and persistence length of a chain projected onto the surface of confinement. Similar to Section 8.2.1, the projected 2D chain theory describes the chain as a two-dimensional walk, making

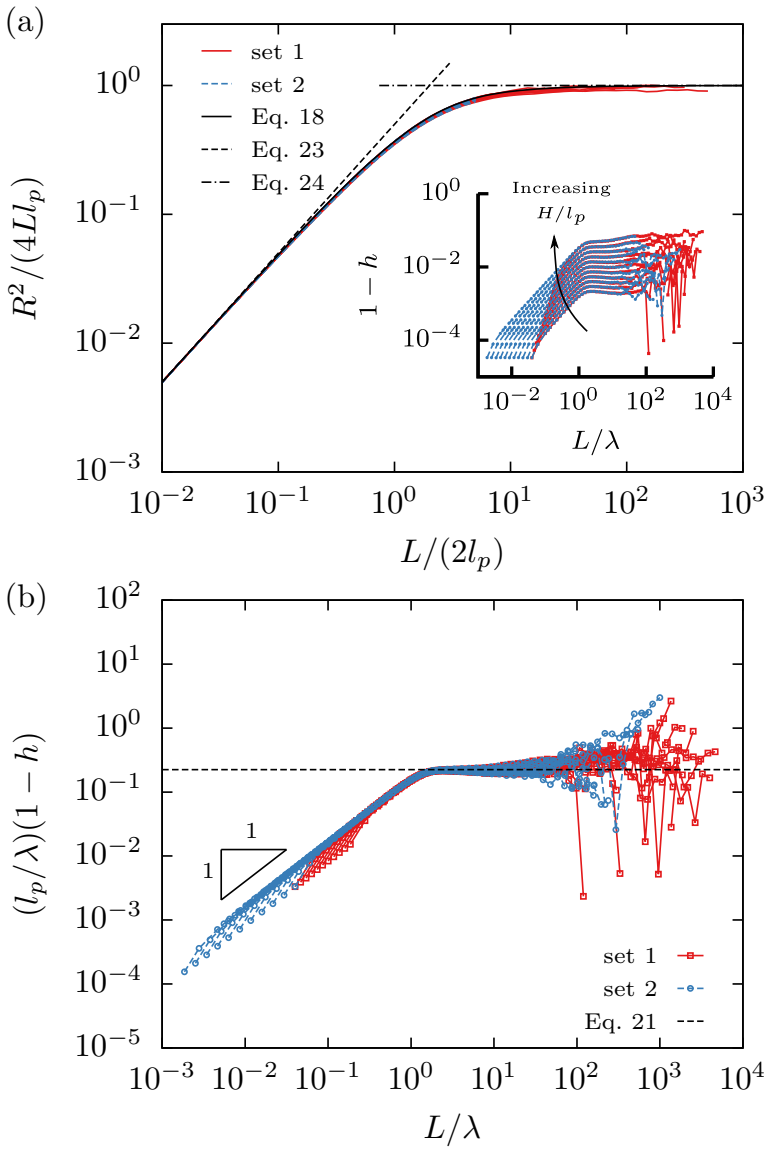

Figure 8.7: Simulation data for the end-to-end distance of ideal chains as a function of contour length. The data in the figure correspond to sets 1 and 2 in Table 8.2. The data in panel (a) is scaled by the size of an ideal, 2D chain. The solid curve in panel (a) corresponds to $g(L/l_p)$ given by Eq. 8.18, the dashed curve gives Eq. 8.23 and the dot-dashed curve gives Eq. 8.24. The inset to panel (a) shows the relative error by approximating $R \approx g$ without including a dependence on the slit height. Panel (b) shows a rescaled value of the relative error which accounts for the first term in the series expansion of $h(\lambda/l_p)$ in Eq. 8.21. The dashed line shows the prediction of Eq. 8.24 with a value of $\alpha_R = 0.1039(5)$.

the difference between the two theories subtle to distinguish. The key difference lies in the dependence of the chain statistics on H/l_p, which is embedded in the parameter $L_\parallel$ (as seen in Figure 8.3) for the projected 2D chain theory and is independent of L for our treatment in Section 8.2.1.

Since we have calculated R^2, $L_\parallel$ and $l_{p,\parallel}$, we are in a position to evaluate the claims of the projected 2D chain theory respecting the dependence of R^2 on H/l_p. According to the theory, the crossovers between subregimes depends on the value of $L_\parallel/l_{p,\parallel}$ rather than L/l_p. Therefore, since $L_\parallel$ depends on H/l_p, one would expect a weak slit height dependence on the crossover between subregimes. However, the data collapse in Figure 8.7(b) does not support this interpretation, since we have assumed that h is independent of L/l_p. Additionally, the projected 2D chain theory predicts that the dependence of R on H/l_p changes between subregimes, because $R \sim L$ for the Odijk-rod subregime and $R \sim L^{1/2}$ for the Odijk-Gaussian subregime. By contrast, the collapse in Figure 8.7(b) shows that the dependence of R on slit height is independent of the change in subregime. Finally, there is a statistically significant (albeit small) difference between the values of the constants α_R and α_L. This difference provides further evidence that the slit height dependence of $L_\parallel$ cannot completely account for the slit height dependence of R.

8.3.3 Real Chains

Building on our discussion of ideal chains, we now turn our attention to the results of PERM simulations which incorporate excluded volume effects. As Section 8.2.2 indicates, one expects the impact of excluded volume to be minimal on the behavior of short chains and on metrics which depend solely on proximal chain-chain interactions. In agreement with this hypothesis, we do not observe changes to $L_\parallel$ following the inclusion of excluded volume interactions, and we therefore omit it from further discussion. However, we do expect a change in behavior for metrics such as the end-to-end distance and confinement free energy, which depend on interactions between distal polymer segments. Such expectations were outlined in Section 8.2.2, where we harmonized a Flory theory by Odijk [136] and the theory for ideal chains in Section 8.2.1.

One of the most apparent predictions of the Flory theory in Section 8.2.1 is a change in the scaling of the end-to-end distance from $R \sim L^{1/2}$ to $R \sim L^{3/4}$ for self-avoiding

chains in the long-chain limit. To test this prediction, Figure 8.8(a) shows the computed mean square end-to-end distance (normalized by the size of an ideal chain) as a function of the contour length (normalized by $2l_p$). Note that this normalization changes the slope to $R^2/L \sim L^0$ for ideal chains and to $R^2/L \sim L^{1/2}$ for swollen chains. In the figure, the end-to-end distance is calculated at a fixed value of H/l_p, while w/H is free to vary. As such, the change in scaling of the polymer size as the excluded volume strength varies is unambiguous, since the degree of freedom due to w/H is absent for ideal chains.

The results shown in Figure 8.8(a) agree with Odijk's Flory theory and do indeed give a slope of 1/2 in the long chain limit. This is most readily apparent for chains where $w \to H$, which show a very short crossover to the self-avoiding behavior. We further anticipate that chains with $w \ll H$ also show a slope of 1/2 in the long chain limit, but finite computational resources limit our ability to check this hypothesis. In addition to agreement with the slope predicted by scaling theory, the results agree with Eq. 8.35, describing the end-to-end distance in the so-called Odijk-Flory subregime. Using a least-squares fit to the data in the figure, we obtain the prefactor $c = 0.593(3)$ for Eq. 8.35, which agrees with our expectation that c should be O(1).

Beyond describing the long-chain limit, Odijk's theory also predicts the crossover behavior according to our original conceptualization of the problem embodied in the schematic in Figure 8.1(b). The results in Figure 8.8(a) are in good agreement with this prediction. For chains with $w \ll H$, the behavior very closely resembles the ideal chain case from Figure 8.7(a) and follows the curve describing the end-to-end distance of an ideal, two-dimensional wormlike chain given by Eq. 8.18. As w/H increases, the excluded volume strength increases and the curve shows a slope of 1/2, indicating a two-dimensional self-avoiding walk. Figure 8.8(b) further examines the crossover in terms of the excluded volume parameter z given in Eq. 8.32. The data collapse nicely to a single curve in the large z limit, which confirms that Eq. 8.32 does indeed describe the crossover to self-avoiding chains in the Odijk regime.

In Figure 8.1 we also proposed that a strongly confined polymer should behave like a two-dimensional, self-avoiding wormlike chain in the limit that $w \to H$. To evaluate this claim, Figure 8.8(a) and Figure 8.8(b) also show data from lattice simulations of 2D, self-avoiding wormlike chain from Hsu *et. al.* [208]. To compare both data sets, it was necessary to compute the persistence length of the lattice chains simulated by Hsu

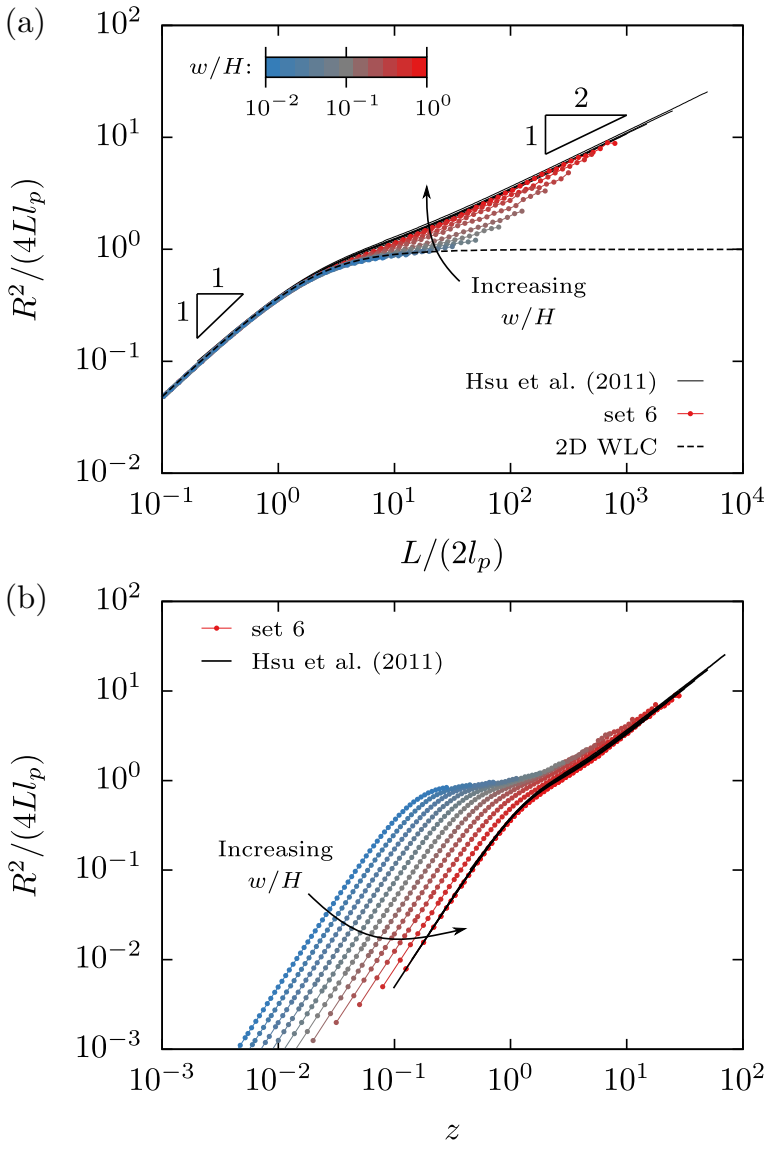

Figure 8.8: PERM data for the normalized end-to-end distance of real chains as a function of (a) normalized projected contour length and (b) excluded volume parameter. In both panels, simulations are run at a fixed value of $H/l_p = 10^{1.5}$, and the different colored symbols correspond to varying values of w/l_p as indicated by the colorbar. PERM data correspond to set 6 in Table 8.2. The solid lines in both panels (a) and (b) correspond to data found in Hsu *et. al.* [208] for 2D wormlike chains on a lattice.

et. al. in a manner consistent with our approach from Section 8.3.1. Therefore, instead of using the empirically determined values given in Ref. 208, we used the relation

$$l_p = l_b \left[2 \exp\left(-\frac{\epsilon_b}{k_B T} \right) \right]^{-1} \tag{8.46}$$

where l_b is the bond length of the lattice chain and ϵ_b is the energy penalty for bending. Using this definition of the persistence length and the data for the R^2 versus L in Ref. 208, we find excellent agreement with the slit data in the limit that $w \to H$. In fact, with no fitting parameters, both the scaling crossover and prefactors agree, providing strong evidence that as the chain width approaches H, the slit-confined chain becomes two-dimensional.

As was the case with ideal chains, we expect that the polymer size will have a weak dependence on the change in deflection segment length as the slit height changes in the form of some function $\tilde{h}(\lambda/l_p)$. In the ideal chain case, we had a closed form expression for $g(L/l_p)$, so we could examine h by the decomposition

$$h = \frac{R^2}{l_p^2} \frac{1}{g} \tag{8.47}$$

which is equivalent to its definition in Eq. 8.17. However, while $g(L/l_p)$ is known for ideal chains, no closed form expression is available for $\tilde{g}(L/l_p, w/H)$ when excluded volume is introduced. We circumvent this problem by performing multiple simulations at constant L/l_p and w/H and look for the relative change in R^2 when the value of λ/l_p is varied.

Defining the relative change in the square end-to-end distance as the difference between R^2 at one value of λ/l_p, and R_0^2 at another, much smaller value of $\lambda_0/l_{p,0}$, we obtain

$$\frac{R_0^2 - R^2}{R_0^2} = \frac{\tilde{g}_0 \tilde{h}_0 - \tilde{g}\tilde{h}}{\tilde{g}_0 \tilde{h}_0} \tag{8.48}$$

In Eq. 8.48, $\tilde{g}$ and $\tilde{g}_0$ cancel as long as we hold L/l_p and w/H constant, which is easily done in PERM. Assuming that $\tilde{h}$ can be expanded like Eq. 8.21, and dropping terms of order $(\lambda/l_p)^2$ gives

$$1 - \frac{R^2}{R_0^2} = 2\alpha_R \left(\frac{\lambda}{l_p} - \frac{\lambda_0}{l_{p,0}} \right) \tag{8.49}$$

Note that when $\lambda_0/l_{p,0} \to 0$, $R_0 \to l_p^2 \tilde{g}$ and we recover the relative error expression used in Section 8.3.2.

We proceed to examine $\tilde{h}$ in Figure 8.9 in a manner similar to our analysis of the ideal function h in Figure 8.7. Accordingly, the inset shows the relative change in the end-to-end distance, $1 - R^2/R_0^2$, as a function of the number of deflection segments, which does indeed change as we vary H/l_p. Similar to the inset in Figure 8.7(a), the change is small but systematic. To collapse the data, Eq. 8.49 is rearranged to rescale the y-axis by dividing by the difference in λ/l_p and the result is shown in the main panel in Figure 8.9. The data collapse very cleanly for small values of L/λ, but show significant sampling error for larger contour lengths, similar to Figure 8.7(b). The axis choice in Fig. 8.9 exaggerates the sampling error, since the relative error between the two measures of R^2 is divided by the difference between two small numbers: λ/l_p and $\lambda_0/l_{p,0}$. A fit to the data in the range $L \in [1, 10]$ gives a value of the constant $\alpha_R = 0.108(1)$, which is close to the value of 0.1039(5) for ideal chains obtained in Section 8.3.2. Indeed the agreement is good, considering that the error estimates given are from fits over limited ranges due to the sampling error for large L. The fact that α_R is similar between real chains and ideal chains parallels the consistency of $L_{\|}$, and provides evidence that the formation of deflection segments is unaffected by the inclusion of excluded volume interactions.

Finally we close this section with a discussion of the free energy of the confined chain. Odijk's theory in Section 8.2.2 presents two components that contribute to the total free energy of a confined chain with excluded volume relative to an ideal, unconfined reference state

$$\Delta F_{\text{conf,ev}} = \Delta F_c + \Delta F_{\text{ex}} \tag{8.50}$$

The first is the free energy required to confine an ideal chain, ΔF_c, and the second is the free energy due to excluded volume interactions of the confined chain, ΔF_{ex}. As it often appears in the literature, it is also convenient to define a free energy of confinement $\Delta \tilde{F}_c$ as the free energy required to confine a chain with excluded volume interactions relative to an excluded volume reference state.

The main panel in Figure 8.10 shows an example of all three of these free energies (normalized by $k_B T$) as a function of the number of persistence lengths. Two obvious conclusions jump out from the data. First, we observe that $\Delta F_c \gg \Delta F_{\text{ex}}$, and therefore the free energy required to confine a chain dominates the total free energy. This finding matches our intuition developed from Odijk's Flory theory in Section 8.2.2. Second, we see that $\Delta F_c \approx \Delta \tilde{F}_c$. This observation is related to the fact that the excess free energy

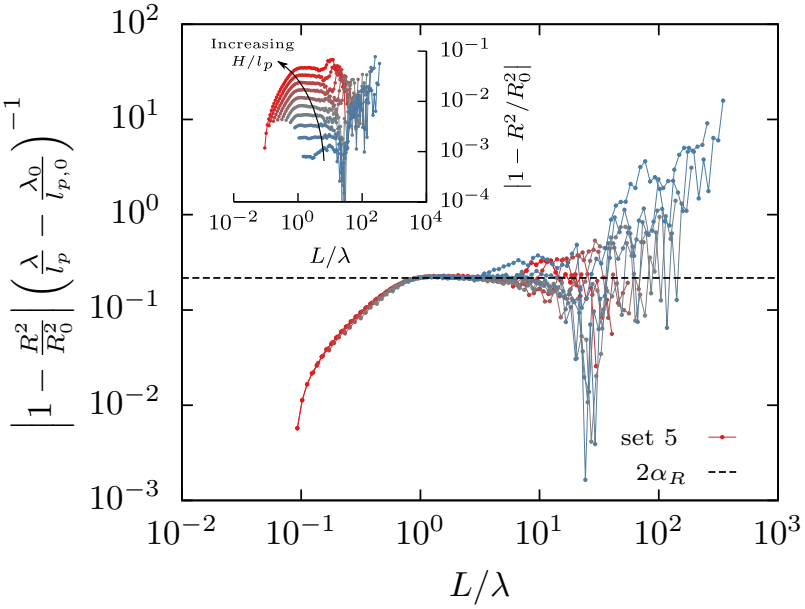

Figure 8.9: The relative change in square end-to-end distance between simulations with different values of H/l_p as a function of the number of deflection segments. The inset shows the absolute value of the relative change and the main panel shows the absolute value of the relative change scaled by the difference in λ/l_p between the simulations. The absolute value is used since the data are presented on a log scale and sampling error leads to a fluctuation in the sign of the error for large L. The data in the figure were obtained at a constant value of $H/w = 5$ and correspond to set 5 in Table 8.2. The dashed line in the figure corresponds to a value of 2α. The value of H/l_p varies from 10^{-3} to 10^{-1}.

is much smaller than the confinement free energy, but we stress that the statements are not equivalent since $\Delta\tilde{F}_c$ has a different reference state. Nevertheless, since the two confinement free energies are so similar, for most practical purposes the confinement free energy of a chain with excluded volume can be described by Eq. 8.12, which was originally discussed in terms of the ideal chain only.

While not dominant in terms of magnitude, the excess free energy is still important, since it determines the crossover from Odijk-Gaussian to Odijk-Flory behavior. Thus for small values of L/l_p the excess free energy is practically zero as indicated by the noisy fluctuations about the value of 10^{-4} for $\beta\Delta F_{\text{ex}}$ in Figure 8.10. However as L/l_p increases, the number of possible intrachain interactions follows suit, and the excess free energy eventually scales linearly in L/l_p, parallel to ΔF_c in the long chain limit. The linear scaling in L/l_p contradicts the Flory theory in Section 8.2.2, but coincides with the predictions of modern renormalization group theory [78].

The excess free energy also changes with excluded volume strength, but since Flory theory fails to describe this effect, we are not aware of any theory that makes a quantitative prediction to which we can compare. Accordingly, we simply report the change in the excess free energy per persistence length for very long chains as a function of w/H in the inset in Figure 8.10. In the figure we observe that the excess free energy is small for small w, monotonically increases as the ratio w/H gets larger and saturates as w approaches the slit height.

Before leaving the subject of confined chains with excluded volume, we comment briefly on previous work that has also posited the existence of self-crossing and non-self-crossing regimes for strongly confined chains in slits [142]. As the nomenclature implies, the distinction between these possible regimes occurs when $H = 2w$; when $H < 2w$, excluded volume interactions prevent chain segments from crossing, whereas when $H \geq 2w$ no such restriction is imposed. In contrast with this idea, Figures 8.8 and 8.10 show a gradual excluded volume crossover suggested by Eq. 8.30 rather than a topologically constrained state. Due to these observations, we find no evidence to support the idea that a prohibition of self-crossing is responsible for the self-avoidance of strongly confined chains in slits. Furthermore, we suggest that a potential crossover at $H = 2w$ may not be universal and would be an artifact of hard beads, since softer beads could cross, depending on the choice of potential and length scale.

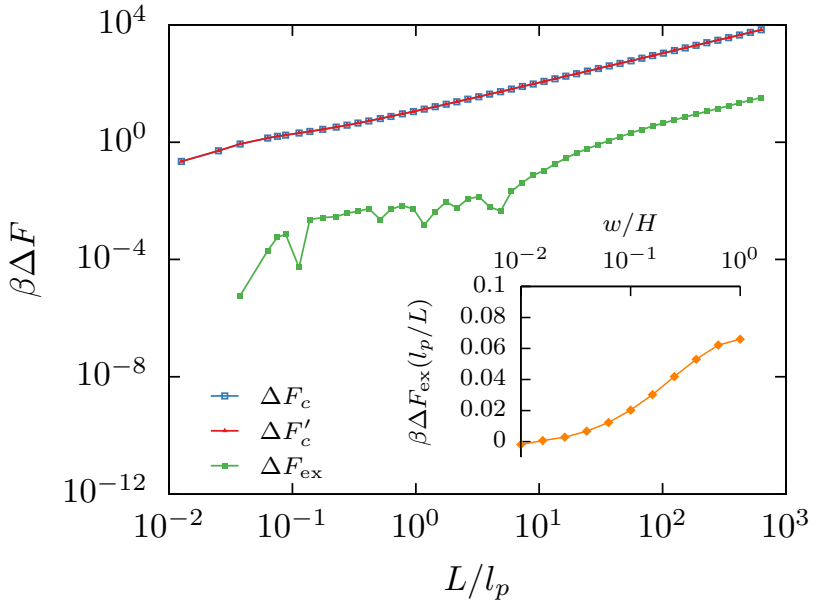

Figure 8.10: (main) Normalized ideal free energy of confinement, real free energy of confinement and excess free energy versus the number of persistence lengths for a chain with $H/l_p = 10^{-1.5}$ and $w/H = 0.398$. (inset) Excess free energy per persistence length for long chains ($L/a = 50001$) with a fixed value of $H/l_p = 10^{-1.5}$ and varying values of w/H. The data used for both the main panel and the inset correspond to sets 3, 5 and 7 in Table 8.2.

8.4 Conclusion

We have presented a comprehensive theory for the Odijk regime in nanoslits that reconciles details from both the literature on ideal chains and the literature on real chains. In particular we have shown that there are three regimes of behavior for slits where $H \ll l_p$, which we labelled the weak confinement regime, the orientational regime and the Odijk regime. Furthermore, we demonstrated three subregimes of behavior for long chains: an Odijk-rod regime, and Odijk-Gaussian regime and an Odijk-Flory regime, the last of which only occurs for chains with excluded volume. For the Odijk regime in particular we developed a theory which describes the end-to-end distance in terms of a two-dimensional wormlike chain with out-of-plane fluctuations described by a series expansion in λ/l_p.

Numerical results are in excellent agreement with the theory. Results for the projected persistence length and confinement free energy are consistent with the three regime picture, and results for the end-to-end distance provide firm evidence for three subregimes within the Odijk regime. Additionally, we find quantitative agreement with the theory for the end-to-end distance, and using our PERM simulations we find several previously unknown prefactors for chains with excluded volume. Interestingly we find that, while conceptually important, the influence of the slit height on the end-to-end distance of an ideal chain is practically negligible. However, this is not the case for real chains, where the slit height strongly influences the excluded volume interactions of the chain.

Finally we note that this problem is quite mature in terms of theory and simulation, yet work remains to relate these rather abstract regimes to actual experimental systems. Such experiments can be quite difficult in practice, since one must use a system with a large persistence length $l_p \gg H$ and long chains $L \gg H$ relative to the slit height. Despite the difficulty, there are notable examples of recent successes in measuring semiflexible chains in the Odijk regime by Nöding and Köster [134] and by Frykholm *et. al.* [135]. Nevertheless, it remains to be seen if the regimes we have outlined can be observed in real systems, or whether effects such as soft or long-range potentials play a more dominant role.

Chapter 9

Conclusion

From the outset, we stated our desire to create a numerical model capable of producing engineering quality results for designing nanochannel genomic mapping devices. As a complementary goal, we also set out to explore and understand the basic physics driving the stretching (or perhaps more appropriately the compression) of DNA in nanochannels. We have made progress on both fronts and several interesting challenges remain.

The project has been quite successful from a methodological perspective, and we have adapted new tools to the problem. In particular, the off-lattice implementation of PERM has been extraordinarily useful for accessing long chains that were unreachable by Markov chain Monte Carlo techniques. It is anticipated that further innovations are possible with PERM, since the technique is relatively new and its failure mechanism is only understood at a rudimentary level. However, it is unclear at this point whether an order of magnitude increase in accessible chain length would produce key insights, or simply be an algorithmic exercise.

In Chapter 4, we examined properties of dilute solutions of DNA in the light of the wormlike chain model and the universal properties of polymers. Interestingly, we found that DNA converges particularly slowly to the long chain limit that one expects from standard polymer physics. From a polymer physics perspective, the statistical mechanics of a dilute solution of wormlike chains is a mature problem, and it appears that there is little room for further fundamental study. Nevertheless, the exact details of the crossover of the fully fluctuating diffusivity from rod-like behavior to Zimm-like behavior

has yet to be fully understood [278] and is surely an exception to this statement. This problem is especially difficult because of a lack of methods for simultaneously dealing with both long-chain hydrodynamics and long-chain excluded volume interactions. As such, most studies focus on either the Kirkwood diffusivity (as we have), or on dynamic simulations of very short chains with Brownian dynamics. In addition, even with well-known fundamentals, several issues remain in mapping the parameters of the wormlike chain model to dsDNA. In particular the role of intercalating dyes and electrostatics appear to be very important and are not very well understood.

Chapter 5 explored the possibility of a "Gauss–de Gennes" regime due to weak excluded volume interactions at intermediate channel sizes. In this context, the use of PERM enabled the exploration of significantly longer chain lengths and the calculation of free energies. However, very recent work by Dai *et. al.* [276] suggests that our results may be better interpreted as evidence for the "Extended de Gennes", due to subtle second-order effects in the free energy. If so, the origin of the "-1" scaling in the extension near $D \approx l_p$ remains a mystery. Indeed, by their interpretation, it is not clear if this observed scaling is a universal regime or a coincidence from a specific polymer model. This problem remains important, since as we have repeated several times, practically all of the nanochannel mapping occurs in or around the "-1" regime, where the physics is complicated. Indeed, regardless of the minor distinction between these regimes, it appears that intermediate regimes are responsible for the observed deviation from de Gennes scaling by Reisner *et. al.* [60].

While the equilibrium behavior of DNA in nanochannels seems to be sharpening in focus, the dynamic behavior remains largely unexplored. We have begun this work and offer estimates of the diffusivity and relaxation time of DNA in nanochannels in Chapters 6 and 7. Nevertheless, in both cases we estimate the dynamic properties using Metropolis Monte Carlo and the Kirkwood diffusivity in only a limited range of the parameter space. More sophisticated methods, such as PERM offer the possibility to greatly expand this range. Furthermore, as in free solution, there are legitimate questions about the accuracy of the Kirkwood approximation, but these remain very challenging to address. Indeed, while it appears that Brownian Dynamics simulations of long chains including hydrodynamic interactions could address this problem, simulating long enough chains have already proved a difficult challenge for Monte Carlo algorithms.

In Chapter 8 we explored the Odijk regime in slits and showed that Odijk's intuition of a random walk of deflection segments could be written into a more quantitative theory. In addition, this chapter demonstrates the interesting strengths and weaknesses of nanochannels and nanoslits. It appears that true slits are likely only useful for studying the physics of confined DNA, since fractional extension decreases with increasing channel size. However, from an experimental perspective, they may be an ideal platform due to their ease of fabrication. It is also quite interesting to explore the analogies between the confinement in a nanochannel versus confinement in a slit. While much of this work has been done for the de Gennes regime using blob theory, investigation into other regimes may provide a wealth of insight. Furthermore, the cross-over from slit-like confinement to channel confinement has yet to be systematically explored with the accompanying focus on the role of rectangular nanochannels that are so prevalently used in experiments.

Finally, we conclude by discussing some of the opportunities and challenges in applying our numerical model directly to genomic mapping problems. After much effort, we appear to have a well characterized and parameterized model for DNA, which has already provided insights into the genomic mapping process. For instance, in Chapter 7, we showed that the relaxation time drops precipitously when the channel size is smaller than the persistence length of the chain, which is key to the design of genomic mapping devices. While this type of information is useful and provides a needed theoretical justification for empirical practices, we hope our model can be more useful as a predictive tool, rather than simply a vehicle for *ex post* storytelling. We made one such prediction in Chapter 6, where we stated that the diffusivity of DNA near $D \approx l_p$ should be independent of the blob size — a result which has yet to be directly tested experimentally. However, we hope that more practical insights can be gleaned from the model and leveraged in an effort to design future genomic mapping devices. Indeed, to be truly useful, we hope that our numerical model can be used as a tool for innovation for the next generation of genomic mapping technology.

Bibliography

[1] I. H. G. S. Consortium. "Initial sequencing and analysis of the human genome". *Nature* 409 (2001), 860–921.

[2] I. H. G. S. Consortium. "Finishing the euchromatic sequence of the human genome". *Nature* 431 (2004), 931–945.

[3] P. Ball. "DNA: Celebrate the unknowns". *Nature* 496 (2013), 419–420.

[4] F. S. Collins, E. D. Green, A. E. Guttmacher, and M. S. Guyer. "A vision for the future of genomics research". *Nature* 422 (2003), 835–847.

[5] A. Balmain, J. Gray, and B. Ponder. "The genetics and genomics of cancer". *Nat. Genet.* 33 (2003), 238–244.

[6] C. R. Marshall et al. "Structural Variation of Chromosomes in Autism Spectrum Disorder". *The American Journal of Human Genetics* 82 (2008), 477 –488.

[7] M. Snyder, J. Du, and M. Gerstein. "Personal genome sequencing: current approaches and challenges". *Gene. Dev.* 24 (2010), 423–431.

[8] S. Zhou et al. "Validation of rice genome sequence by optical mapping". *BMC Genomics* 8 (2007), 278.

[9] S. Zhou et al. "A Single Molecule Scaffold for the Maize Genome". *Plos Genet.* 5 (2009), e1000711.

[10] A. R. Hastie et al. "Rapid Genome Mapping in Nanochannel Arrays for Highly Complete and Accurate De Novo Sequence Assembly of the Complex Aegilops tauschii Genome". *PLoS ONE* 8 (2013), e55864.

[11] B. L. Karger and A. Guttman. "DNA sequencing by CE". *Electrophoresis* 30 (2009), S196–S202.

[12] M. L. Metzker. “Sequencing technologies – the next generation”. *Nat Rev Genet* 11 (2010), 31–46.

[13] T. P. Niedringhaus, D. Milanova, M. B. Kerby, M. P. Snyder, and A. E. Barron. “Landscape of Next-Generation Sequencing Technologies”. *Anal. Chem.* 83 (2011), 4327–4341.

[14] J. R. Miller, S. Koren, and G. Sutton. “Assembly algorithms for next-generation sequencing data”. *Genomics* 95 (2010), 315–327.

[15] E. Mardis, J. McPherson, R. Martienssen, R. K. Wilson, and W. R. McCombie. “What is Finished, and Why Does it Matter”. *Genome Res.* 12 (2002), 669–671.

[16] S. Zhou et al. “A Whole-Genome Shotgun Optical Map of Yersinia pestis Strain KIM”. *Appl. Environ. Microbiol* 68 (2002), 6321–6331.

[17] M. Margulies et al. “Genome sequencing in microfabricated high-density picolitre reactors”. *Nature* 437 (2005), 376–380.

[18] M. R. Speicher and N. P. Carter. “The new cytogenetics: blurring the boundaries with molecular biology”. *Nat. Rev. Genet.* 6 (2005), 782–792.

[19] D. M. Hinckley, G. S. Freeman, J. K. Whitmer, and J. J. de Pablo. “An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization”. *J. Chem. Phys.* 139 (2013), 144903.

[20] K. A. Wetterstrand. *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program.* 2014. URL: `https://www.genome.gov/sequencingcosts/`.

[21] R. M. Durbin. “A map of human genome variation from population-scale sequencing”. *Nature* 467 (2010), 1061–1073.

[22] D. Schwartz, X. Li, L. Hernandez, S. Ramnarain, E. Huff, and Y. Wang. “Ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping”. *Science* 262 (1993), 110–114.

[23] L. Feuk, A. R. Carson, and S. W. Scherer. “Structural variation in the human genome”. *Nat. Rev. Genet.* 7 (2006), 85–97.

[24] R. Redon et al. “Global variation in copy number in the human genome”. *Nature* 444 (2006), 444–454.

[25] J. Eid et al. “Real-Time DNA Sequencing from Single Polymerase Molecules”. *Science* 323 (2009), 133–138.

[26] D. Branton et al. “The potential and challenges of nanopore sequencing”. *Nat Biotech* 26 (2008), 1146–1153.

[27] S. Koren, G. Harhay, T. Smith, J. Bono, D. Harhay, S. Mcvey, D. Radune, N. Bergman, and A. Phillippy. “Reducing assembly complexity of microbial genomes with single-molecule sequencing”. *Genome Biology* 14 (2013), R101.

[28] F. Persson and J. O. Tegenfeldt. “DNA in nanochannels-directly visualizing genomic information”. *Chemical Society Reviews* 39 (2010), 985–999.

[29] S. L. Levy and H. G. Craighead. “DNA manipulation, sorting, and mapping in nanofluidic systems”. *Chem. Soc. Rev.* 39 (2010), 1133–1152.

[30] W. Reisner, J. N. Pedersen, and R. H. Austin. “DNA confinement in nanochannels: physics and biological applications”. *Rep. Prog. Phys.* 75 (2012), 106601.

[31] K. D. Dorfman, S. B. King, D. W. Olson, J. D. P. Thomas, and D. R. Tree. “Beyond Gel Electrophoresis: Microfluidic Separations, Fluorescence Burst Analysis, and DNA Stretching”. *Chem. Rev.* 113 (2013), 2584–2667.

[32] M. Levy-Sakin and Y. Ebenstein. “Beyond sequencing: optical mapping of DNA in the age of nanotechnology and nanoscopy”. *Current Opinion in Biotechnology* 24 (2013), 690–698.

[33] M. Levy-Sakin et al. “Toward Single-Molecule Optical Mapping of the Epigenome”. *ACS Nano* 8 (2014), 14–26.

[34] K. D. Dorfman, D. Gupta, A. Jain, and A. Muralidhar. “Hydrodynamics of confined DNA”. *European Physical Journal - Special Topics* invited, to be submitted (2014).

[35] J. Jing et al. “Optical Mapping of Plasmodium falciparum Chromosome 2”. *Genome Res.* 9 (1999), 175–181.

[36] A. Lim et al. “Shotgun optical maps of the whole Escherichia coli O157 : H7 genome”. *Genome Res.* 11 (2001), 1584–1593.

[37] S. Zhou et al. "Whole-Genome Shotgun Optical Mapping of Rhodobacter sphaeroides strain 2.4.1 and Its Use for Whole-Genome Shotgun Sequence Assembly". *Genome Res.* 13 (2003), 2142–2151.

[38] S. Zhou et al. "Shotgun optical mapping of the entire Leishmania major Friedlin genome". *Mol. Biochem. Parasit.* 138 (2004), 97–106.

[39] J. Herrick, X. Michalet, C. Conti, C. Schurra, and A. Bensimon. "Quantifying single gene copy number by measuring fluorescent probe lengths on combed genomic DNA". *Proc. Natl. Acad. Sci. U.S.A.* 97 (2000), 222–227.

[40] J. Giacalone et al. "Optical Mapping of BAC Clones from the Human Y Chromosome DAZ Locus". *Genome Res.* 10 (2000), 1421–1429.

[41] S. Zhou et al. "Single-molecule approach to bacterial genomic comparisons via optical mapping". *J. Bacteriol.* 186 (2004), 7773–7782.

[42] S. Reslewic et al. "Whole-genome shotgun optical mapping of Rhodospirillum rubrum". *Appl Environ Microbiol* 71 (2005), 5511–5522.

[43] P. Latreille et al. "Optical mapping as a routine tool for bacterial genome sequence finishing". *BMC Genomics* 8 (2007), 321.

[44] G. Ananiev et al. "Optical mapping discerns genome wide DNA methylation profiles". *BMC Mol. Biol.* 9 (2008), 68.

[45] C.-w. Wu, T. M. Schramm, S. Zhou, D. C. Schwartz, and A. M. Talaat. "Optical mapping of the Mycobacterium avium subspecies paratuberculosis genome". *BMC Genomics* 10 (2009), 25.

[46] B. Teague et al. "High-resolution human genome structure by single-molecule analysis". *Proc. Natl. Acad. Sci. U.S.A.* 107 (2010), 10848–10853.

[47] E. T. Lam et al. "Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly". *Nat. Biotech.* 30 (2012), 771–776.

[48] X. Meng, K. Benson, K. Chada, E. J. Huff, and D. C. Schwartz. "Optical mapping of lambda bacteriophage clones using restriction endonucleases". *Nat. Genet.* 9 (1995), 432–438.

[49] J. Jing et al. "Automated high resolution optical mapping using arrayed, fluid-fixed DNA molecules". *Proc. Natl. Acad. Sci. USA* 95 (1998), 8046–8051.

[50] X. Hu, C. Aston, and D. C. Schwartz. "Optical Mapping of DNA Polymerase I Action and Products". *Biochem. Bioph. Res. Co.* 254 (1999), 466–473.

[51] K. Jo, D. M. Dhingra, T. Odijk, J. J. de Pablo, M. D. Graham, R. Runnheim, D. Forrest, and D. C. Schwartz. "A single-molecule barcoding system using nanoslits for DNA analysis". *Proc. Natl. Acad. Sci. U.S.A.* 104 (2007), 2673–2678.

[52] A. Valouev, D. C. Schwartz, S. Zhou, and M. S. Waterman. "An algorithm for assembly of ordered restriction maps from single DNA molecules". *Proc. Natl. Acad. Sci. U.S.A.* 103 (2006), 15770–15775.

[53] E. T. Dimalanta et al. "A Microfluidic System for Large DNA Molecule Arrays". *Anal. Chem.* 76 (2004), 5293–5301.

[54] H. Cao, J. O. Tegenfeldt, R. H. Austin, and S. Y. Chou. "Gradient nanostructures for interfacing microfluidics and nanofluidics". *Applied Physics Letters* 81 (2002), 3058–3060.

[55] J. O. Tegenfeldt et al. "The dynamics of genomic-length DNA molecules in 100-nm channels". *Proc. Natl. Acad. Sci. USA* 101 (2004), 10979–10983.

[56] S. K. Das, M. D. Austin, M. C. Akana, P. Deshpande, H. Cao, and M. Xiao. "Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes". *Nucleic Acids Res.* 38 (2010), e177.

[57] R. Riehn, M. Lu, Y. Wang, S. Lim, E. Cox, and R. Austin. "Restriction mapping in nanofluidic devices". *Proc. Natl. Acad. Sci. U.S.A* 102 (2005), 10012–10016.

[58] T. Odijk. "On the statistics and dynamics of confined or entangled stiff polymers". *Macromolecules* 16 (1983), 1340–1344.

[59] M. Daoud and P.-G. de Gennes. "Statistics of macromolecular solutions trapped in small pores". *J. Phys. (Paris)* 38 (1977), 85–93.

[60] W. Reisner et al. "Statics and Dynamics of Single DNA Molecules Confined in Nanochannels". *Phys. Rev. Lett.* 94 (2005), 196101.

[61] A. L. Mackay. *Dictionary of Scientific Quotations*. Institute of Physics Publishing, 1994.

[62] O. Kratky and G. Porod. "X-ray investigation of dissolved chain molecules." *Recl. Trav. Chim. Pays-Bas Belg.* 68 (1949), 1106–1122.

[63] S. Smith, L. Finzi, and C. Bustamante. "Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads". *Science* 258 (1992), 1122–1126.

[64] C. Bustamante, J. Marko, E. Siggia, and S. Smith. "Entropic Elasticity of Lambda-Phage DNA". *Science* 265 (1994), 1599–1600.

[65] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante. "Ionic effects on the elasticity of single DNA molecules". *Proc. Natl. Acad. Sci. USA* 94 (1997), 6185–6190.

[66] C.-C. Hsieh, A. Balducci, and P. S. Doyle. "Ionic Effects on the Equilibrium Dynamics of DNA Confined in Nanoslits". *Nano Lett.* 8 (2008), 1683–1688.

[67] M. C. Linak. "Modeling Complex Structures in Nucleic Acids". PhD thesis. University of Minnesota – Twin Cities, 2012.

[68] N. Saito, K. Takahashi, and Y. Yunoki. "The statistical mechanical theory of stiff chains". *J. Phys. Soc. Jpn.* 22 (1967), 219 –226.

[69] K. F. Freed. *Renormalization Group Theory of Macromolecules*. John Wiley & Sons, Inc., 1987.

[70] G. H. Fredrickson. *The Equilibrium Theory of Inhomogeneous Polymers*. Oxford University Press, 2006.

[71] Y. Wang, D. R. Tree, and K. D. Dorfman. "Simulation of DNA Extension in Nanochannels". *Macromolecules* 44 (2011), 6594–6604.

[72] M. Rubinstein and R. H. Colby. *Polymer Physics*. Oxford University Press, 2003.

[73] H. Benoit and P. Doty. "Light scattering from non-Gaussian chains". *J. Phys. Chem.* 57 (1953), 958–963.

[74] P. C. Hiemenz and T. P. Lodge. *Polymer Chemistry: Second Edition*. CRC Press, 2007.

[75] Y. Wang, W. F. Reinhart, D. R. Tree, and K. D. Dorfman. "Resolution limit for DNA barcodes in the Odijk regime". *Biomicrofluidics* 6 (2012), 014101.

[76] P. J. Flory. *Principles of Polymer Chemistry*. Cornell University Pressr, 1953.

[77] P.-G. de Gennes. *Scaling Concepts in Polymer Physics*. Cornell University Press, 1979.

[78] L. Schäfer. *Excluded Volume Effects in Polymer Solutions as Explained by the Renormalization Group*. Springer-Verlag, 1999.

[79] J. des Cloizeaux. "On the absence of flory terms in the energy and in the entropy of a polymer chain". *J. Phys. France* 37 (1976), 431–434.

[80] B. Li, N. Madras, and A. D. Sokal. "Critical exponents, hyperscaling, and universal amplitude ratios for two- and three-dimensional self-avoiding walks". *J. Stat. Phys.* 80 (1995), 661–754.

[81] N. Clisby. "Accurate Estimate of the Critical Exponent ν for Self-Avoiding Walks via a Fast Implementation of the Pivot Algorithm". *Phys. Rev. Lett.* 104 (2010), 055702.

[82] H. B. Gray Jr. and J. E. Hearst. "Flexibility of native DNA from the sedimentation behavior as a function of molecular weight and temperature". *J. Mol. Biol.* 35 (1968), 111–129.

[83] K. Soda and A. Wada. "Dynamic light-scattering studies on thermal motions of native DNAs in solution". *Biophys. Chem.* 20 (1984), 185–200.

[84] J. E. Godfrey. "The flexibility of low molecular weight double-stranded dna as a function of length: I. Isolation and physical characterization of seven fractions". *Biophys. Chem.* 5 (1976), 285–299.

[85] Z. Kam, N. Borochov, and H. Eisenberg. "Dependence of laser light scattering of DNA on NaCl concentration". *Biopolymers* 20 (1981), 2671–2690.

[86] G. L. Lukacs, P. Haggie, O. Seksek, D. Lechardeur, N. Freedman, and A. S. Verkman. "Size-dependent DNA Mobility in Cytoplasm and Nucleus". *J. Biol. Chem.* 275 (2000), 1625–1629.

[87] G. Voordouw, Z. Kam, N. Borochov, and H. Eisenberg. "Isolation and physical studies of the intact supercoiled: The open circular and the linear forms of CoIE1-plasmid DNA". *Biophys. Chem.* 8 (1978), 171–189.

[88] J. E. Godfrey and H. Eisenberg. "The flexibility of low molecular weight double-stranded dna as a function of length: II. Light scattering measurements and the estimation of persistence lengths from light scattering, sedimentation and viscosity". *Biophys. Chem.* 5 (1976), 301–318.

[89] D. E. Smith, T. T. Perkins, and S. Chu. "Dynamical Scaling of DNA Diffusion Coefficients". *Macromolecules* 29 (1996), 1372–1373.

[90] M. L. Mansfield and J. F. Douglas. "Transport Properties of Wormlike Chains with Applications to Double Helical DNA and Carbon Nanotubes". *Macromolecules* 41 (2008), 5412–5421.

[91] J. Wilcoxon and J. M. Schurr. "Temperature dependence of the dynamic light scattering of linear ϕ29 DNA: Implications for spontaneous opening of the double-helix". *Biopolymers* 22 (1983), 2273–2321.

[92] R. M. Robertson, S. Laib, and D. E. Smith. "Diffusion of isolated DNA molecules: Dependence on length and topology". *Proc. Natl. Acad. Sci. U.S.A* 103 (2006), 7310–7314.

[93] S. S. Sorlie and R. Pecora. "A dynamic light scattering study of a 2311 base pair DNA restriction fragment". *Macromolecules* 21 (1988), 1437–1449.

[94] T. Nicolai and M. Mandel. "Dynamic light scattering by aqueous solutions of low-molar-mass DNA fragments in the presence of sodium chloride". *Macromolecules* 22 (1989), 2348–2356.

[95] J. A. Harpst and J. R. Dawson. "Low angle light scattering studies on whole, half, and quarter molecules of T2 bacteriophage DNA". *Biophys. J.* 55 (1989), 1237–1249.

[96] W. Eimer, J. R. Williamson, S. G. Boxer, and R. Pecora. "Characterization of the overall and internal dynamics of short oligonucleotides by depolarized dynamic light scattering and NMR relaxation measurements". *Biochemistry* 29 (1990), 799–811.

[97] G. F. Bonifacio, T. Brown, G. L. Conn, and A. N. Lane. "Comparison of the electrophoretic and hydrodynamic properties of DNA and RNA oligonucleotide duplexes". *Biophys. J.* 73 (1997), 1532–1538.

[98] A. I. Krasna. “Changes in the light-scattering properties of DNA on denaturation”. *Journal of Colloid and Interface Science* 39 (1972), 632–646.

[99] S. Allison, S. S. Sorlie, and R. Pecora. “Brownian dynamics simulations of wormlike chains: dynamic light scattering from a 2311 base pair DNA fragment”. *Macromolecules* 23 (1990), 1110–1118.

[100] R. Pecora. “DNA: A Model Compound for Solution Studies of Macromolecules”. *Science* 251 (1991), 893–898.

[101] W. Eimer and R. Pecora. “Rotational and translational diffusion of short rodlike molecules in solution: Oligonucleotides”. *J. Chem. Phys.* 94 (1991), 2324–2329.

[102] K. Fukudome, K. Yamaoka, and H. Ochiai. “Ultrasonic Scission of Deoxyribonucleic Acid in Aqueous Solution III. The Solution Properties of Sonicated Low Molecular Weight Samples as Revealed by Light Scattering and Viscosity”. *Polymer Journal* 19 (1987), 1385–1394.

[103] H. Lederer, R. P. May, J. K. Kjems, G. Baer, and H. Heumann. “Solution structure of a short DNA fragment studied by neutron scattering”. *Eur. J. Biochem.* 161 (1986), 191–196.

[104] J. Seils and T. Dorfmüller. “Internal dynamics of linear and superhelical DNA as studied by photon correlation spectroscopy”. *Biopolymers* 31 (1991), 813–825.

[105] R. T. Kovacic and K. E. Van Holde. “Sedimentation of homogeneous double-strand DNA molecules”. *Biochemistry* 16 (1977), 1490–1498.

[106] C. W. Schmid, F. P. Rinehart, and J. E. Hearst. “Statistical length of DNA from light scattering”. *Biopolymers* 10 (1971), 883–893.

[107] D. Jolly and H. Eisenberg. “Photon correlation spectroscopy, total intensity light scattering with laser radiation, and hydrodynamic studies of a well fractionated DNA sample”. *Biopolymers* 15 (1976), 61–95.

[108] S. S. Sorlie and R. Pecora. “A dynamic light scattering study of four DNA restriction fragments”. *Macromolecules* 23 (1990), 487–497.

[109] F. Latinwo and C. M. Schroeder. “Model systems for single molecule polymer dynamics”. *Soft Matter* 7 (2011), 7907–7913.

[110] J. G. Kirkwood and J. Riseman. "The Intrinsic Viscosities and Diffusion Constants of Flexible Macromolecules in Solution". *J. Chem. Phys.* 16 (1948), 565–573.

[111] J. G. Kirkwood. "The general theory of irreversible processes in solutions of macromolecules". *J. Polym. Sci.* 12 (1954), 1–14.

[112] J. J. Erpenbeck and J. G. Kirkwood. "Statistical Mechanics of Irreversible Processes in Polymer Solutions". *J. Chem. Phys.* 29 (1958), 909–913.

[113] J. J. Erpenbeck and J. G. Kirkwood. "Erratum : Statistical Mechanics of Irreversible Processes in Polymer Solutions". *J. Chem. Phys.* 38 (1963), 1023–1024.

[114] M. Doi and S. F. Edwards. *The Theory of Polymer Dynamics*. Oxford University Press, 1986.

[115] R. Zwanzig. "Translational Diffusion in Polymer Solutions". *J. Chem. Phys.* 45 (1966), 1858–1859.

[116] H. Yamakawa. *Modern Theory of Polymer Solutions*. Harper and Row, 1971.

[117] H. C. Ottinger. *Stochastic Processes in Polymeric Fluids: Tools and Examples for Developing Simulation Algorithms*. Springer-Verlag, 1996.

[118] R. B. Bird, C. F. Curtiss, R. C. Armstrong, and O. Hassager. *Dynamics of Polymeric Liquids, Volume 2, Kinetic Theory*. Vol. 2. John Wiley & Sons, Inc., 1986.

[119] R. M. Jendrejack, M. D. Graham, and J. J. de Pablo. "Hydrodynamic interactions in long chain polymers: Application of the Chebyshev polynomial approximation in stochastic simulations". *J. Chem. Phys.* 113 (2000), 2894–2900.

[120] M. Fixman. "Inclusion of hydrodynamic interaction in polymer dynamical simulations". *Macromolecules* 14 (1981), 1710–1717.

[121] B. Liu and B. Dunweg. "Translational diffusion of polymer chains with excluded volume and hydrodynamic interactions by Brownian dynamics simulation". *J. Chem. Phys.* 118 (2003), 8061–8072.

[122] P. Hagerman and B. Zimm. "Monte-Carlo Approach to the Analysis of the Rotational Diffusion of Wormlike Chains". *Biopolymers* 20 (1981), 1481–1502.

[123] M. Fixman. “Symmetry criteria in the theory of chain polymer hydrodynamics”. *Macromolecules* 14 (1981), 1706–1709.

[124] R. Rodríguez Schmidt, J. Hernández Cifre, and J. Garcìa de la Torre. “Translational diffusion coefficients of macromolecules”. *Eur. Phys. J. E* 35 (2012), 1–5.

[125] P. E. Rouse. “A theory of the linear viscoelastic properties of dilute solutions of coiling polymers”. *J. Chem. Phys.* 21 (1953), 1272–1280.

[126] B. H. Zimm. “Dynamics of Polymer Molecules in Dilute Solution: Viscoelasticity, Flow Birefringence and Dielectric Loss”. *J. Chem. Phys.* 24 (1956), 269–278.

[127] Y. Oono and M. Kohmoto. “Renormalization group theory of transport properties of polymer solutions. I. Dilute solutions”. *J. Chem. Phys.* 78 (1983), 520–528.

[128] J. F. Douglas and K. F. Freed. “Competition between Hydrodynamic Screening (”Draining”) and Excluded Volume Interactions in an Isolated Polymer Chain”. *Macromolecules* 27 (1994), 6088–6099.

[129] H. Yamakawa and M. Fujii. “Translational Friction Coefficient of Wormlike Chains”. *Macromolecules* 6 (1973), 407–415.

[130] F. Brochard and P.-G. De Gennes. “Dynamics of confined polymer-chains”. *J. Chem. Phys.* 67 (1977), 52–56.

[131] Y. Yang, T. W. Burkhardt, and G. Gompper. “Free energy and extension of a semiflexible polymer in cylindrical confining geometries”. *Phys. Rev. E* 76 (2007), 011804.

[132] T. W. Burkhardt, Y. Yang, and G. Gompper. “Fluctuations of a long, semiflexible polymer in a narrow channel”. *Phys. Rev. E* 82 (2010), 041801.

[133] F. Persson, P. Utko, W. Reisner, N. B. Larsen, and A. Kristensen. “Confinement Spectroscopy: Probing Single DNA Molecules with Tapered Nanochannels”. *Nano Lett.* 9 (2009), 1382–1385.

[134] B. Nöding and S. Köster. “Intermediate Filaments in Small Configuration Spaces”. *Phys. Rev. Lett.* 108 (2012), 088101.

[135] K. Frykholm, M. Alizadehheidari, J. Fritzsche, J. Wigenius, M. Modesti, F. Persson, and F. Westerlund. "Probing Physical Properties of a DNA-Protein Complex Using Nanofluidic Channels". *Small* 10 (2014), 884–887.

[136] T. Odijk. "Scaling theory of DNA confined in nanochannels and nanoslits". *Phys. Rev. E* 77 (2008), 060901(R).

[137] F. Brochard-Wyart and E. Raphael. "Scaling theory of molten polymers in small pores". *Macromolecules* 23 (1990), 2276–2280.

[138] F. Brochard-Wyart, T. Tanaka, N. Borghi, and P.-G. de Gennes. "Semiflexible Polymers Confined in Soft Tubes". *Langmuir* 21 (2005), 4144–4148.

[139] P. Cifra. "Channel confinement of flexible and semiflexible macromolecules". *J. Chem. Phys.* 131 (2009), 224903.

[140] P. Cifra. "Weak-to-strong confinement transition of semi-flexible macromolecules in slit and in channel". *J. Chem. Phys.* 136 (2012), 024902.

[141] Y.-L. Chen, M. D. Graham, J. J. de Pablo, G. C. Randall, M. Gupta, and P. S. Doyle. "Conformation and dynamics of single DNA molecules in parallel-plate slit microchannels". *Phys. Rev. E* 70 (2004), 060901(R).

[142] L. Dai, J. J. Jones, J. R. C. van der Maarel, and P. S. Doyle. "A systematic study of DNA conformation in slitlike confinement". *Soft Matter* 8 (2012), 2972–2982.

[143] P.-K. Lin, C. Fu, Y. Chen, Y. Chen, P. Wei, C. Kuan, and W. Fann. "Static conformation and dynamics of single DNA molecules confined in nanoslits". *Phys. Rev. E.* 76 (2007), 011806.

[144] T. Sakaue and E. Raphael. "Polymer Chains in Confined Spaces and Flow-Injection Problems: Some Remarks". *Macromolecules* 39 (2006), 2621–2628.

[145] Y. Wang, I. Teraoka, F. Y. Hansen, G. H. Peters, and O. Hassager. "Mean Span Dimensions of Ideal Polymer Chains Containing Branches and Rings". *Macromolecules* 44 (2011), 403–412.

[146] D. J. Bonthuis, C. Meyer, D. Stein, and C. Dekker. "Conformation and Dynamics of DNA Confined in Slitlike Nanofluidic Channels". *Phys. Rev. Lett.* 101 (2008), 108303.

[147] J. Tang, S. L. Levy, D. W. Trahan, J. J. Jones, H. G. Craighead, and P. S. Doyle. "Revisiting the conformation and dynamics of DNA in slitlike confinement". *Macromolecules* 43 (2010), 7368–7377.

[148] E. A. Strychalski, J. Geist, M. Gaitan, L. E. Locascio, and S. M. Stavis. "Quantitative Measurements of the Size Scaling of Linear and Circular DNA in Nanofluidic Slitlike Confinement". *Macromolecules* 45 (2012), 1602–1611.

[149] K. Lin, C.-C. Hsieh, Y.-L. Chen, and C.-F. Chou. "Effects of Topology and Ionic Strength on Double-Stranded DNA Confined in Nanoslits". *Macromolecules* 45 (2012), 2920–2927.

[150] P. Cifra, Z. Benkova, and T. Bleha. "Persistence Lengths and Structure Factors of Wormlike Polymers under Confinement". *J. Phys. Chem. B* 112 (2008), 1367–1375.

[151] P. Cifra, Z. Benkova, and T. Bleha. "Effect of confinement on properties of stiff biological macromolecules". *Faraday Discuss.* 139 (2008), 377–392.

[152] A. Balducci, P. Mao, J. Han, and P. S. Doyle. "Double-stranded DNA diffusion in slitlike nanochannels". *Macromolecules* 39 (2006), 6273–6281.

[153] E. A. Strychalski, S. L. Levy, and H. G. Craighead. "Diffusion of DNA in nanoslits". *Macromolecules* 41 (2008), 7716–7721.

[154] P.-K. Lin, J.-F. Chang, C.-H. Wei, P. H. Tsao, W. S. Fann, and Y.-L. Chen. "Partial hydrodynamic screening of confined linear and circular double-stranded DNA dynamics". *Phys. Rev. E* 84 (2011), 031917.

[155] L. Dai, D. R. Tree, J. R. C. van der Maarel, K. D. Dorfman, and P. S. Doyle. "Revisiting Blob Theory for DNA Diffusivity in Slitlike Confinement". *Phys. Rev. Lett.* 110 (2013), 168105.

[156] A. Balducci, C.-C. Hsieh, and P. Doyle. "Relaxation of Stretched DNA in Slitlike Confinement". *Phys. Rev. Lett.* 99 (2007), 238102.

[157] C.-C. Hsieh, A. Balducci, and P. S. Doyle. "An Experimental Study of DNA Rotational Relaxation Time in Nanoslits". *Macromolecules* 40 (2007), 5196–5205.

[158] J. Tang, D. W. Trahan, and P. S. Doyle. "Coil–Stretch Transition of DNA Molecules in Slitlike Confinement". *Macromolecules* 43 (2010), 3081–3089.

[159] D. W. Trahan and P. S. Doyle. "Simulating the Relaxation of Stretched DNA in Slitlike Confinement". *Macromolecules* 44 (2011), 383–392.

[160] J. I. Siepmann. "A method for the direct calculation of chemical potentials for dense chain systems". *Mol. Phys.* 6 (1990), 1145–1158.

[161] D. Frenkel and B. Smit. *Understanding Molecular Simulation from Algorithms to Applications*. Academic Press, 2002.

[162] J. A. Gubner. *Probability and Random Processes for Electrical and Computer Engineers*. Oxford University Press, 2006.

[163] F. T. Wall and F. Mandel. "Macromolecular dimensions obtained by an efficient Monte Carlo method without sample attrition". *J. Chem. Phys.* 63 (1975), 4592–4595.

[164] S. K. Kumar, M. Vacatello, and D. Y. Yoon. "Off-lattice Monte Carlo simulations of polymer melts confined between two plates". *J. Chem. Phys.* 89 (1988), 5206–5215.

[165] N. Madras and A. D. Sokal. "The pivot algorithm: A highly efficient Monte Carlo method for the self-avoiding walk". *J. Stat. Phys.* 50 (1988), 109–186.

[166] P. H. Verdier and W. H. Stockmayer. "Monte Carlo Calculations on the Dynamics of Polymers in Dilute Solution". *J. Chem. Phys.* 36 (1962), 227–235.

[167] T. Prellberg. "From Rosenbluth Sampling to PERM - rare event sampling with stochastic growth algorithms" (2012), 311–334. arXiv:1205.3734v1.

[168] M. N. Rosenbluth and A. W. Rosenbluth. "Monte Carlo Calculation of the Average Extension of Molecular Chains". *J. Chem. Phys.* 23 (1955), 356–359.

[169] H. Meirovitch. "A new method for simulation of real chains: scanning future steps". *J. Phys. A: Math. Gen.* 15 (1982), L735.

[170] F. T. Wall and J. J. Erpenbeck. "New Method for the Statistical Computation of Polymer Dimensions". *J. Chem. Phys.* 30 (1959), 634–637.

[171] P. Grassberger. "Pruned-enriched Rosenbluth method: Simulations of theta polymers of chain length up to 1,000,000". *Phys. Rev. E* 56 (1997), 3682–3693.

[172] H.-P. Hsu, W. Paul, and K. Binder. "Polymer chain stiffness vs. excluded volume: A Monte Carlo study of the crossover towards the worm-like chain model". *Europhys. Lett.* 92 (2010), 28003.

[173] T. Prellberg and J. Krawczyk. "Flat Histogram Version of the Pruned and Enriched Rosenbluth Method". *Phys. Rev. Lett.* 92 (2004), 120602.

[174] J. R. Blake. "A note on the image system for a stokeslet in a no-slip boundary". *Proc. Camb. Phil. Soc.* 70 (1971), 303–310.

[175] N. Liron and S. Mochon. "Stokes flow for a stokeslet between two parallel flat plates". *J. Eng. Math.* 10 (1976), 287–303.

[176] R. M. Jendrejack, D. C. Schwartz, M. D. Graham, and J. J. de Pablo. "Effect of confinement on DNA dynamics in microfluidic devices". *J. Chem. Phys.* 119 (2003), 1165–1173.

[177] C. Pozrikidis. *Introduction to Theoretical and Computational Fluid Dynamics*. Oxford University Press, 1997.

[178] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in Fortran: The Art of Scientific Computing*. 2001.

[179] C. Brockman, S. J. Kim, and C. M. Schroeder. "Direct observation of single flexible polymers using single stranded DNA". *Soft Matter* 7 (2011), 8005–8012.

[180] M. Nepal, A. Yaniv, E. Shafran, and O. Krichevsky. "Structure of DNA Coils in Dilute and Semidilute Solutions". *Phys. Rev. Lett.* 110 (2013), 058102.

[181] J. A. Harpst. "Analysis of low angle light scattering results from T7 DNA". *Biophys. Chem.* 11 (1980), 295 –302.

[182] C. M. Schroeder, E. S. G. Shaqfeh, and S. Chu. "Effect of Hydrodynamic Interactions on DNA Dynamics in Extensional Flow: Simulation and Single Molecule Experiment". *Macromolecules* 37 (2004), 9242–9256.

[183] M. L. Mansfield and J. F. Douglas. "Influence of variable hydrodynamic interaction strength on the transport properties of coiled polymers". *Phys. Rev. E* 81 (2010), 021803.

[184] O. B. Bakajin, T. A. J. Duke, C. F. Chou, S. S. Chan, R. H. Austin, and E. C. Cox. "Electrohydrodynamic stretching of DNA in confined environments". *Phys. Rev. Lett.* 80 (1998), 2737–2740.

[185] A. Sischka, K. Toensing, R. Eckel, S. D. Wilking, N. Sewald, R. Ros, and D. Anselmetti. "Molecular Mechanisms and Kinetics between DNA and DNA Binding Ligands". *Biophys. J.* 88 (2005), 404–411.

[186] K. Günther, M. Mertig, and R. Seidel. "Mechanical and structural properties of YOYO-1 complexed DNA". *Nucleic Acids Res.* 38 (2010), 6526–6532.

[187] H.-P. Hsu, W. Paul, and K. Binder. "Scattering function of semiflexible polymer chains under good solvent conditions". *J. Chem. Phys.* 137 (2012), 174902.

[188] D. Amorós, A. Ortega, and J. García de la Torre. "Hydrodynamic Properties of Wormlike Macromolecules: Monte Carlo Simulation and Global Analysis of Experimental Data". *Macromolecules* 44 (2011), 5788–5797.

[189] C. Storm and P. C. Nelson. "Theory of high-force DNA stretching and overstretching". *Phys. Rev. E* 67 (2003), 051906.

[190] J. Wang and H. Gao. "A generalized bead-rod model for Brownian dynamics simulations of wormlike chains under strong confinement". *J. Chem. Phys.* 123 (2005), 084906.

[191] R. M. Jendrejack, J. J. de Pablo, and M. D. Graham. "Stochastic simulations of DNA in flow: Dynamics and the effects of hydrodynamic interactions". *J. Chem. Phys.* 116 (2002), 7752–7759.

[192] J. F. Marko and E. D. Siggia. "Stretching DNA". *Macromolecules* 28 (1995), 8759–8770.

[193] S. B. Smith, Y. Cui, and C. Bustamante. "Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules". *Science* 271 (1996), 795–799.

[194] O. A. Saleh, D. B. McIntosh, P. Pincus, and N. Ribeck. "Nonlinear Low-Force Elasticity of Single-Stranded DNA Molecules". *Phys. Rev. Lett.* 102 (2009), 068301.

[195] A. Y. L. Sim, J. Lipfert, D. Herschlag, and S. Doniach. "Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle x-ray scattering". *Phys. Rev. E* 86 (2012), 021901.

[196] H. Chen, S. P. Meisburger, S. A. Pabit, J. L. Sutton, W. W. Webb, and L. Pollack. "Ionic strength-dependent persistence lengths of single-stranded RNA and DNA". *Proc. Natl. Acad. Sci. USA* 109 (2012), 799–804.

[197] M. C. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, and T. Ha. "Probing Single-Stranded DNA Conformational Flexibility Using Fluorescence Spectroscopy". *Biophys. J.* 86 (2004), 2530–2537.

[198] J. A. Schellman. "Flexibility of DNA". *Biopolymers* 13 (1974), 217–226.

[199] H. Jian, A. V. Vologodskii, and T. Schlick. "A Combined Wormlike-Chain and Bead Model for Dynamic Simulations of Long Linear DNA". *J Comput. Phys.* 136 (1997), 168 –179.

[200] H. Frauenkron, M. S. Causo, and P. Grassberger. "Two-dimensional self-avoiding walks on a cylinder". *Phys. Rev. E* 59 (1999), R16–R19.

[201] B. J. Cherayil, J. F. Douglas, and K. F. Freed. "Change of a flexible polymer's free energy due to excluded volume, molecular architecture and the presence of boundaries". *Macromolecules* 20 (1987), 1345–1353.

[202] J. S. Pedersen, M. Laso, and P. Schurtenberger. "Monte Carlo study of excluded volume effects in wormlike micelles and semiflexible polymers". *Phys. Rev. E* 54 (1996), R5917–R5920.

[203] B. H. Zimm. "Chain Molecule Hydrodynamics by the Monte-Carlo Method and the Validity of the Kirkwood-Riseman Approximation". *Macromolecules* 13 (1980), 592–602.

[204] J. E. Hearst and W. H. Stockmayer. "Sedimentation Constants of Broken Chains and Wormlike Coils". *J. Chem. Phys.* 37 (1962), 1425–1433.

[205] H. C. Öttinger. "Translational diffusivity from the Zimm model". *J. Chem. Phys.* 87 (1987), 3156–3165.

[206] H.-P. Hsu and K. Binder. "Stretching semiflexible polymer chains: Evidence for the importance of excluded volume effects from Monte Carlo simulation". *J. Chem. Phys.* 136 (2012), 024901.

[207] R. Radhakrishnan and P. T. Underhill. "Models of flexible polymers in good solvents: relaxation and coil-stretch transition". *Soft Matter* 8 (2012), 6991–7003.

[208] H.-P. Hsu, W. Paul, and K. Binder. "Breakdown of the Kratky-Porod wormlike chain model for semiflexible polymers in two dimensions". *Europhys. Lett.* 95 (2011), 68004.

[209] J. Moon and H. Nakanishi. "Onset of the excluded-volume effect for the statistics of stiff chains". *Phys. Rev. A* 44 (1991), 6427–6442.

[210] A. Y. Grosberg and A. R. Khokhlov. *Statistical Physics of Macromolecules.* American Institute of Physics, 1994.

[211] M. Muthukumar and B. G. Nickel. "Expansion of a polymer chain with excluded volume interaction". *J. Chem. Phys.* 86 (1987), 460–476.

[212] Z. Y. Chen and J. Noolandi. "Renormalization-group scaling theory for flexible and wormlike polymer chains". *J. Chem. Phys.* 96 (1992), 1540–1548.

[213] N. Borochov, H. Eisenberg, and Z. Kam. "Dependence of DNA conformation on the concentration of salt". *Biopolymers* 20 (1981), 231–235.

[214] D. Porschke. "Persistence length and bending dynamics of DNA from electrooptical measurements at high salt concentrations". *Biophys. Chem.* 40 (1991), 169–179.

[215] W. F. Reed, S. Ghosh, G. Medjahdi, and J. Francois. "Dependence of polyelectrolyte apparent persistence lengths, viscosity, and diffusion on ionic strength and linear charge density". *Macromolecules* 24 (1991), 6189–6198.

[216] T. Odijk. "Polyelectrolytes near the rod limit". *Journal of Polymer Science: Polymer Physics Edition* 15 (1977), 477–483.

[217] J. Skolnick and M. Fixman. "Electrostatic Persistence Length of a Wormlike Polyelectrolyte". *Macromolecules* 10 (1977), 944–948.

[218] A. V. Dobrynin. "Effect of Counterion Condensation on Rigidity of Semiflexible Polyelectrolytes". *Macromolecules* 39 (2006), 9519–9527.

[219] D. Stigter. "Interactions of Highly Charged Colloidal Cylinders with Applications to Double Stranded DNA". *Biopolymers* 16 (1977), 1435–1448.

[220] V. V. Rybenkov, N. R. Cozzarelli, and A. V. Vologodskii. "Probability of DNA knotting and the effective diameter of the DNA double helix". *Proc. Natl. Acad. Sci. USA* 90 (1993), 5307–5311.

[221] D. B. McIntosh, N. Ribeck, and O. A. Saleh. "Detailed scaling analysis of low-force polyelectrolyte elasticity". *Phys. Rev. E* 80 (2009), 041803.

[222] B. Tinland, A. Pluen, J. Sturm, and G. Weill. "Persistence Length of Single-Stranded DNA". *Macromolecules* 30 (1997), 5763–5765.

[223] M. C. Linak, R. Tourdot, and K. D. Dorfman. "Moving beyond Watson–Crick models of coarse grained DNA dynamics". *J. Chem. Phys.* 135 (2011), 205102.

[224] A. J. Spakowitz and Z.-G. Wang. "Exact Results for a Semiflexible Polymer Chain in an Aligning Field". *Macromolecules* 37 (2004), 5814–5823.

[225] J. S. Pedersen and P. Schurtenberger. "Scattering Functions of Semiflexible Polymers with and without Excluded Volume Effects". *Macromolecules* 29 (1996), 7602–7612.

[226] P. Sharp and V. A. Bloomfield. "Light scattering from wormlike chains with excluded volume effects". *Biopolymers* 6 (1968), 1201–1211.

[227] G. S. Manning. "Limiting laws and counterion condensation in polyelectrolyte solutions. 7. Electrophoretic mobility and conductance". *The Journal of Physical Chemistry* 85 (1981), 1506–1515.

[228] I. M. Withers, A. V. Dobrynin, M. L. Berkowitz, and M. Rubinstein. "Monte Carlo simulation of homopolymer chains. I. Second virial coefficient". *J. Chem. Phys.* 118 (2003), 4721–4732.

[229] H. Yamakawa. "Stiff-Chain Macromolecules". *Annu. Rev. Phys. Chem.* 35 (1984), 23–47.

[230] N. C. Stellwagen, S. Magnusdottir, C. Gelfi, and P. G. Righetti. "Measuring the translational diffusion coefficients of small DNA molecules by capillary electrophoresis". *Biopolymers* 58 (2001), 390–397.

[231] Y. Lu, B. Weers, and N. C. Stellwagen. "DNA persistence length revisited". *Biopolymers* 61 (2002), 261–275.

[232] H. S. Rye, S. Yue, D. E. Wemmer, M. A. Quesada, R. P. Haugland, R. A. Mathies, and A. N. Glazer. "Stable fluorescent complexes of double-stranded DNA with bis-intercalating asymmetric cyanine dyes: properties and applications". *Nucleic Acids Res.* 20 (1992), 2803–2812.

[233] C. M. Schroeder, H. P. Babcock, E. S. G. Shaqfeh, and S. Chu. "Observation of Polymer Conformation Hysteresis in Extensional Flow". *Science* 301 (2003), 1515–1519.

[234] J. W. Halley, D. Atkatz, and H. Nakanishi. "Scaling at the rod-to-flexible chain crossover in the stiff limit". *J. Phys. A* 23 (1990), 3297.

[235] Z. Dogic et al. "Elongation and Fluctuations of Semiflexible Polymers in a Nematic Solvent". *Phys. Rev. Lett.* 92 (2004), 125503.

[236] E. F. Casassa. "Comments on Exclusion of Polymer Chains from Small Pores and Its Relation to Gel Permeation Chromatography". *Macromolecules* 9 (1976), 182–185.

[237] E. Werner, F. Persson, F. Westerlund, J. O. Tegenfeldt, and B. Mehlig. "Orientational correlations in confined DNA". *Phys. Rev. E* 86 (2012), 041802.

[238] P. Cifra, Z. Benková, and T. Bleha. "Chain Extension of DNA Confined in Channels". *J. Phys. Chem. B* 113.7 (2009), 1843–1851.

[239] Y. Wang, G. H. Peters, F. Y. Hansen, and O. Hassager. "Equilibrium partitioning of macromolecules in confining geometries: Improved universality with a new molecular size parameter". *J. Chem. Phys.* 128 (2008).

[240] J. Z. Y. Chen and D. E. Sullivan. "Free Energy of a Wormlike Polymer Chain Confined in a Slit: Crossover between Two Scaling Regimes". *Macromolecules* 39 (2006), 7769–7773.

[241] T. Sakaue. "Semiflexible Polymer Confined in Closed Spaces". *Macromolecules* 40 (2007), 5206–5211.

[242] S. Jun, D. Thirumalai, and B.-Y. Ha. "Compression and Stretching of a Self-Avoiding Chain in Cylindrical Nanopores". *Phys. Rev. Lett.* 101 (2008), 138101.

[243] T. Su, S. K. Das, M. Xiao, and P. K. Purohit. "Transition between Two Regimes Describing Internal Fluctuation of DNA in a Nanochannel". *PLoS ONE* 6 (2011), e16890.

[244] C.-B. Kim, H. Chun, J. Chung, K. H. Lee, J. H. Lee, B. Song, and S.-H. Lee. "In Situ Curing of Sliding SU-8 Droplet over a Microcontact Printed Pattern for Tunable Fabrication of a Polydimethylsiloxane Nanoslit". *Anal. Chem.* 83 (2011), 7221–7226.

[245] S. F. Lim, A. Karpusenko, J. J. Sakon, J. A. Hook, T. A. Lamar, and R. Riehn. "DNA methylation profiling in nanochannels". *Biomicrofluidics* 5 (2011), 034106.

[246] F. Wagner, G. Lattanzi, and E. Frey. "Conformations of confined biopolymers". *Phys. Rev. E* 75 (2007), 050902.

[247] J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill. "Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations". *J. Chem. Theory Comput.* 3 (2007), 26–41.

[248] D. C. Morse. "Viscoelasticity of Concentrated Isotropic Solutions of Semiflexible Polymers. 2. Linear Response". *Macromolecules* 31 (1998), 7044–7067.

[249] J. García de la Torre, A. Jiménez, and J. J. Freire. "Monte Carlo calculation of hydrodynamic properties of freely jointed, freely rotating, and real polymethylene chains". *Macromolecules* 15 (1982), 148–154.

[250] J. L. Harden and M. Doi. "Diffusion of macromolecules in narrow capillaries". *J. Phys. Chem.* 96 (1992), 4046–4052.

[251] C. Zhang et al. "Amplified stretch of bottlebrush-coated DNA in nanofluidic channels". *Nucleic Acids Res.* (2013).

[252] R. Chang and K. Jo. "DNA conformation in nanochannels: Monte Carlo simulation studies using a primitive DNA model". *J. Chem. Phys.* 136 (2012), 095101.

[253] W. F. Reinhart, D. R. Tree, and K. D. Dorfman. "Entropic depletion of DNA in triangular nanochannels". *Biomicrofluidics* 7 (2013), 024102.

[254] T. Odijk. "DNA confined in nanochannels: Hairpin tightening by entropic depletion". *J. Chem. Phys.* 125, 204904 (2006), 204904.

[255] C. Zhang, F. Zhang, J. A. van Kan, and J. R. C. van der Maarel. "Effects of electrostatic screening on the conformation of single DNA molecules confined in a nanochannel". *J. Chem. Phys.* 128 (2008), 225109.

[256] Z. Benková and P. Cifra. "Simulation of Semiflexible Cyclic and Linear Chains Moderately and Strongly Confined in Nanochannels". *Macromolecules* 45 (2012), 2597–2608.

[257] C. Manneschi, E. Angeli, T. Ala-Nissila, L. Repetto, G. Firpo, and U. Valbusa. "Conformations of DNA in triangular nanochannels". *Macromolecules* 46 (2013), 4198–4206.

[258] C. Micheletti and E. Orlandini. "Knotting and metric scaling properties of DNA confined in nano-channels: a Monte Carlo study". *Soft Matter* 8 (2012), 10959–10968.

[259] E. Werner, F. Westerlund, J. O. Tegenfeldt, and B. Mehlig. "Monomer Distributions and Intrachain Collisions of a Polymer Confined to a Channel". *Macromolecules* 46 (2013), 6644–6650.

[260] M. D. Graham. "Fluid Dynamics of Dissolved Polymer Molecules in Confined Geometries". *Annu. Rev. Fluid Mech.* 43 (2011), 273–298.

[261] E. F. Koslover and A. J. Spakowitz. "Systematic coarse-graining of microscale polymer models as effective elastic chains". *Macromolecules* 46 (2013), 2003–2014.

[262] T. Cui, J. Ding, and J. Z. Y. Chen. "Dynamics of a self-avoiding polymer chain in slit, tube, and cube confinements". *Phys. Rev. E* 78 (2008).

[263] A. Karpusenko, J. H. Carpenter, C. Zhou, S. F. Lim, J. Pan, and R. Riehn. "Fluctuation modes of nanoconfined DNA". *J. Appl. Phys.* 111 (2012), 024701.

[264] T. C. B. McLeish. "Tube theory of entangled polymer dynamics". *Adv. Phys.* 51 (2002), 1379–1527.

[265] E. F. Casassa. "Equilibrium distribution of flexible polymer chains between macroscopic solution phase and small voids". *J. Polym. Sci., Part B: Polym. Lett.* 5 (1967), 773–778.

[266] S. L. Levy, J. T. Mannion, J. Cheng, C. H. Reccius, and H. G. Craighead. "Entropic Unfolding of DNA Molecules in Nanofluidic Channels". *Nano Lett.* 8 (2008), 3839–3844.

[267] J. Mannion, C. Reccius, J. Cross, and H. Craighead. "Conformational Analysis of Single DNA Molecules Undergoing Entropically Induced Motion in Nanochannels". *Biophysical Journal* 90 (2006), 4538 –4545.

[268] K. D. Dorfman. "DNA electrophoresis in microfabricated devices". *Rev. Mod. Phys.* 82 (2010), 2903–2947.

[269] W. Reisner, J. P. Beech, N. B. Larsen, H. Flyvbjerg, A. Kristensen, and J. O. Tegenfeldt. "Nanoconfinement-Enhanced Conformational Response of Single DNA Molecules to Changes in Ionic Environment". *Phys. Rev. Lett.* 99 (2007), 058302.

[270] K. L. Kounovsky-Shafer, J. P. Hernández-Ortiz, K. Jo, T. Odijk, J. J. de Pablo, and D. C. Schwartz. "Presentation of Large DNA Molecules for Analysis as Nanoconfined Dumbbells". *Macromolecules* 46 (2013), 8356–5368.

[271] B. J. Bicout and T. W. Burkhardt. "Simulation of a semiflexible polymer in a narrow cylindrical pore". *J. Phys. A: Math. Gen.* 34 (2001), 5745–5750.

[272] C. Rivetti, M. Guthold, and C. Bustamante. "Scanning Force Microscopy of DNA Deposited onto Mica: Equilibration versus Kinetic Trapping Studied by Statistical Polymer Chain Analysis". *J. Mol. Biol.* 264 (1996), 919–932.

[273] T. Odijk. "Theory of Lyotropic Polymer Liquid Crystals". *Macromolecules* 19 (1986), 2313–2329.

[274] L. Auvray. "Solutions de macromolécules rigides : effets de paroi, de confinement et d'orientation par un écoulement". *J. Phys. France* 42 (1981), 79–95.

[275] A. Muralidhar, D. R. Tree, Y. Wang, and K. D. Dorfman. "Interplay between chain stiffness and excluded volume of semiflexible polymers confined in nanochannels". *J. Chem. Phys.* 140 (2014), 084905.

[276] L. Dai, J. R. C. van der Maarel, and P. S. Doyle. "Extended de Gennes regime of DNA confined in a Nanochannel". *Macromolecules* 47 (2014), 2445–2450.

[277] H.-P. Hsu, W. Paul, and K. Binder. "Standard Definitions of Persistence Length Do Not Describe the Local "Intrinsic" Stiffness of Real Polymer Chains". *Macromolecules* 43 (2010), 3094–3102.

[278] M. L. Mansfield and J. F. Douglas. "Is duplex DNA a swollen random coil?" *Soft Matter* 9 (2013), 8914–8922.

[279] H. E. Daniels. "XXI.—The Statistical Theory of Stiff Chains". *Proc. R. Soc. Edinb. A* 63 (1952), 290–311.

Appendix A

Supporting Information to Chapter 4

A.1 DWLC Bond Probability Density

It is helpful to describe bond j in local spherical coordinates with the $j-1$ bond endpoint located at and aligned with the origin and the end point of j at $\{\rho_j, \theta_j, \phi_j\}$ where ρ_j is the radial coordinate, θ_j the polar angle and ϕ_j the azimuthal angle. The coordinates are independent so the joint probability density function is

$$P(\rho_j, \theta_j, \phi_j) = f(\rho_j)g(\theta_j)h(\phi_j) \tag{A.1}$$

The bond vectors are inextensible so a simple delta function describes its probability density

$$f(\rho_j) = \delta(\rho_j - a) \tag{A.2}$$

The equilibrium distribution of the polar angle θ_j is proportional to the Boltzmann factor $P \sim \exp[-\kappa(1 - \cos\theta_j)]$. By normalizing, one obtains [198]

$$g(\theta_j) = \frac{\kappa \exp\left[-\kappa(1 - \cos\theta_j)\right]}{1 - \exp(-2\kappa)} \tag{A.3}$$

There is no such constraint on the azimuthal angle, since there is no potential, so its distribution is uniform

$$h(\phi_j) = \frac{1}{2\pi} \tag{A.4}$$

To arrive at Eq. 2 in the main text, we adopt the definition of the Kuhn length as the bond length of an equivalent freely jointed chain [72, 199]

$$\frac{b}{a} = \frac{1 + \langle \cos\theta_j \rangle}{1 - \langle \cos\theta_j \rangle} \tag{A.5}$$

Performing the expectation using Eq. A.3 gives Eq. 2 in the main text.

Eq. A.3 is useful for understanding the role of discretization in the model. Due to the obvious trade-off between increased resolution and computational cost, one may be tempted to treat the bond length as an artifact of a discrete model alone (similar to a time step in a dynamic integration). Indeed, it is precisely for this reason — a choice of model length scale — that we turned to a coarse-grained model in the first place. However, polymer chains (unlike time) are not necessarily best described by continuous models and short chain behavior may also include effects of discrete bond lengths (*i.e.* ssDNA [189, 193–195]). In other words a is a legitimate coarse-graining variable not simply a numerical artifact.

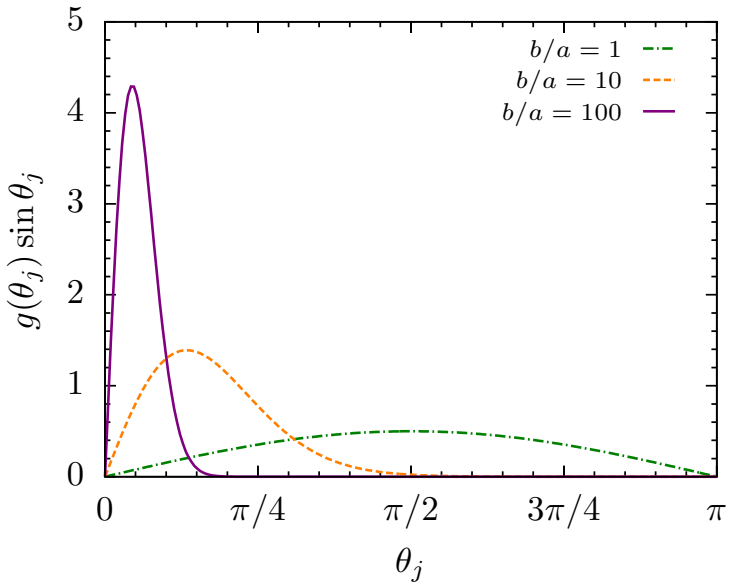

Figure A.1: Effect of the bond length a on the bond angle flexibility distribution function $g(\theta_j)$. When a is small, the bonds are stiffer and the chain exhibits more continuous, "persistent" behavior characteristic of the CWLC. When a is large, the bonds rotate isotropically and the chain is discrete and freely-jointed.

Fig. A.1 shows the effect of the bond length a on the bond distribution function. When the bond length is small relative to the Kuhn length of the chain, the probability distribution is sharp and the chain is indeed "persistent". This corresponds to the CWLC which is recovered when $\{a \to 0, N \to \infty, L \to aN\}$. When $a \sim b$ however, the distribution is broad and isotropic, indicating a freely jointed angle. This corresponds to the case in which $\kappa \to 0$, where we recover the FJC.

A.2 PERM Properties in $O(N)$ Time

Since large chains are used in our PERM calculations, $O(N^2)$ quantities can dominate the running time if they are not carefully accounted for. This is true for any quantity that involves a double-sum over the number of beads, including post-processing calculations (which must be performed at run-time). In particular we find that the excluded volume interactions, Kirkwood diffusivity, radius of gyration and form factor all require $O(N^2)$ time using conventional formulas. In the section that follows we list methods for handling the excluded volume, radius of gyration and form factor in $O(N)$ time. We do not have a method for calculating the Kirkwood diffusivity in $O(N)$ time, which is difficult because hydrodynamic interactions are long-ranged. To make such a calculation in less than $O(N^2)$ time, we anticipate that advanced techniques for long-ranged interactions (such as Ewald or particle-particle/particle-mesh sums) would be required.

To handle excluded volume interactions, we implement a neighbor list. Since PERM is a chain growth method, the neighbor list is somewhat different than in typical Markov-chain Monte Carlo algorithms, so we briefly describe it here. As in all neighbor lists, we seek to avoid a costly search of the ij pairs, which in this case corresponds to beads that are nearby in real space, but distant along the contour of the chain. To begin, we define a cutoff distance, r_c, and at some point as the chain grows, the end-to-end distance exceeds the cut-off. At this point we place the last bead at the center of a sphere of size r_c and compile a list of neighbors. Upon resuming growth, as long as the chain remains inside $r_c - w$ (where w is the short range distance of interaction), we only need to check the neighbors on the list. Of course, as the chain grows the neighbor list must also grow, since these beads lie within the cutoff radius. When chain growth exceeds the cutoff radius, we recompile a neighbor list and re-center the cutoff distance.

In this method, the choice of r_c is critical for the efficiency of the algorithm. If it is too small, the neighbor list is computed too often and the method is effectively $O(N^2)$; if it is too large, the neighbor list grows too long and the method becomes $O(N^2)$ again.

In addition to the excluded volume interactions, post-processing calculations like the radius of gyration and form factor involve double-sums over the number of beads in the chain. Two key steps are necessary for converting a post-processing metric to an $O(N)$ method. First, the formula is converted into an iterative form, which depends on the metric at the previous step. Second, this iterative formula must be evaluated in $O(1)$ time for each chain growth step. The radius of gyration for step n with position, $\mathbf{r}_n$, is given by

$$S_n{}^2 = \overline{R^2}_n - R^2_{cm,n} \tag{A.6}$$

where

$$\overline{R^2}_n = \frac{n-1}{n}\,\overline{R^2}_{n-1} + \frac{r_n{}^2}{n} \tag{A.7}$$

is the mean square position and

$$\mathbf{R}_{cm,n} = \frac{n-1}{n}\,\mathbf{R}_{cm,n-1} + \frac{\mathbf{r}_n}{n} \tag{A.8}$$

is the center of mass vector at step n. In these equations, $r_n^2 = \mathbf{r}_n \cdot \mathbf{r}_n$ is the square of the position vector and $R^2_{cm,n} = \mathbf{R}_{cm,n} \cdot \mathbf{R}_{cm,n}$ is the square center of mass. The form factor for step n with wave vector, $\mathbf{q}$, is given by [187]

$$P^{(n)}(\mathbf{q}) = \frac{a_n{}^2 + b_n{}^2}{(N+1)^2} \tag{A.9}$$

$$a_n = a_{n-1} + \sin\left(\mathbf{q} \cdot \mathbf{r}_n\right) \tag{A.10}$$

$$b_n = b_{n-1} + \cos\left(\mathbf{q} \cdot \mathbf{r}_n\right) \tag{A.11}$$

Since the form factor in free solution is isotropic, the direction of the wave vector is arbitrary. In our case, we fixed the wave vector to be parallel to the x-axis.

A.3 Renormalization Group Theory for WLCs

Chen and Noolandi [212] give a set of coupled, non-linear equations that can be solved to obtain the end-to-end distance and radius of gyration of a wormlike chain with excluded

volume as a function of chain length. The end-to-end distance is given by

$$\frac{R^2}{b^2} = \frac{L\mathcal{L}_R}{b} - \frac{1}{2}\left[1 - \exp\left(-2\frac{L\mathcal{L}_R}{b}\right)\right] \tag{A.12}$$

$$1 - (1-\overline{u})\,\mathcal{L}_R^{-2.624} = \overline{u}\sqrt{1 + 0.34934\frac{R^2}{b^2}}\mathcal{L}_R^{-3.322} \tag{A.13}$$

where

$$\overline{u} = \frac{1}{0.1777}\left(\frac{3}{2\pi}\right)^{3/2}\frac{w_{\mathrm{RG}}}{b} \tag{A.14}$$

is a dimensionless excluded volume term with w_{RG} as the effective width for the renormalization group theory. Also appearing is the rescaled function $\mathcal{L}_R$, which must be solved concurrently with R. The radius of gyration is given by

$$\frac{S^2}{b^2} = \frac{L\mathcal{L}_S}{6b} - \frac{1}{4}\left\{1 - \frac{b}{L\mathcal{L}_S} + \frac{b^2}{2(L\mathcal{L}_S)^2}\left[1 - \exp\left(-2\frac{L\mathcal{L}_S}{b}\right)\right]\right\} \tag{A.15}$$

$$1 - (1-\overline{u})\,\mathcal{L}_S^{-2.624} = \overline{u}\sqrt{1 + 1.6728\frac{S^2}{b^2}}\mathcal{L}_S^{-3.322} \tag{A.16}$$

Again, the scaling function $\mathcal{L}_S$ which appears, must be solved simultaneously with S.

Since the DWLC model and the continuous WLC model implement the excluded volume differently [212], it is unclear how the effective width w of the discrete model matches the effective width of the continuous model w_{RG}. To ensure that both models are consistent, Fig. A.2 shows RG and PERM calculations scaled such that both agree in the limit that $z \to \infty$. Note that when plotted as a function of z, curves for both ssDNA and dsDNA collapse in the high-z limit. A value of $w_{\mathrm{RG}} = 1.50w$ causes the PERM and RG curves to collapse and gives excellent agreement for all values z.

A.4 Wormlike Chain Form Factor

The analytical expression for the form factor of an ideal wormlike chain is given by

$$P(q) = \frac{2}{\mathcal{N}^2}\mathcal{L}^{-1}\left(\frac{G(Q;p)}{p^2}\right), \tag{A.17}$$

where $\mathcal{N}$ is the number of Kuhn monomers, q is the wave vector, p is the Laplace variable and $\mathcal{L}^{-1}$ represents an inverse Laplace transform [224]. $G(Q;p)$ is the three-dimensional

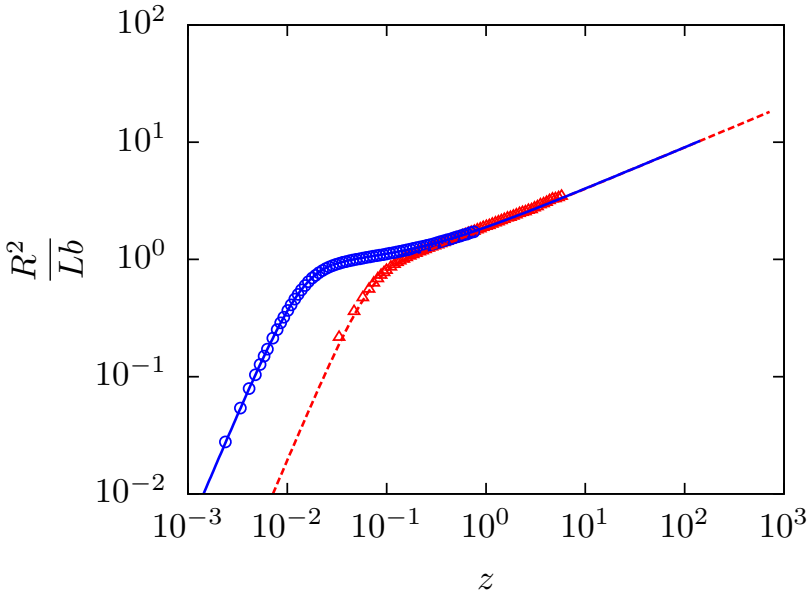

Figure A.2: Normalized square end-to-end distance of a wormlike chain as a function of the excluded volume parameter z. Symbols correspond to PERM simulations (dsDNA is given by open circles and ssDNA is given by open triangles) and lines correspond to RG calculations. The excluded volume parameter in the RG calculations has been rescaled to match the hard core excluded volume in the PERM calculations, $w_{\mathrm{RG}} = 1.50w$.

end-to-end distance distribution function in Fourier-Laplace space and is written in the form of a continued fraction as

$$G(Q;p) = \cfrac{1}{P_0 + \cfrac{a_1^2 Q^2}{P_1 + \cfrac{a_2^2 Q^2}{P_2 + \cfrac{a_3^2 Q^2}{\cdots}}}}. \tag{A.18}$$

Here $Q = 2l_p q$, $P_m = p + m(m+1)$ and $a_m = m/\sqrt{4m^2 - 1}$ with an index m ranging from 0 to ∞. To avoid an infinite summation, the continued fraction in Eq. A.18 was truncated at $m = 20$ as in Spakowitz and Wang [224] to obtain a rational polynomial. The inverse Laplace transform in Eq. A.17 was computed by evaluating the residues at the poles of the latter to recover $P(q)$. In order to verify the PERM calculations, Fig. A.3 compares the form factor of ideal wormlike chains (WLC) from simulations and the analytical results given by Eq. A.17.

In addition, Sharp and Bloomfield [226] give a semi-empirical equation for the form

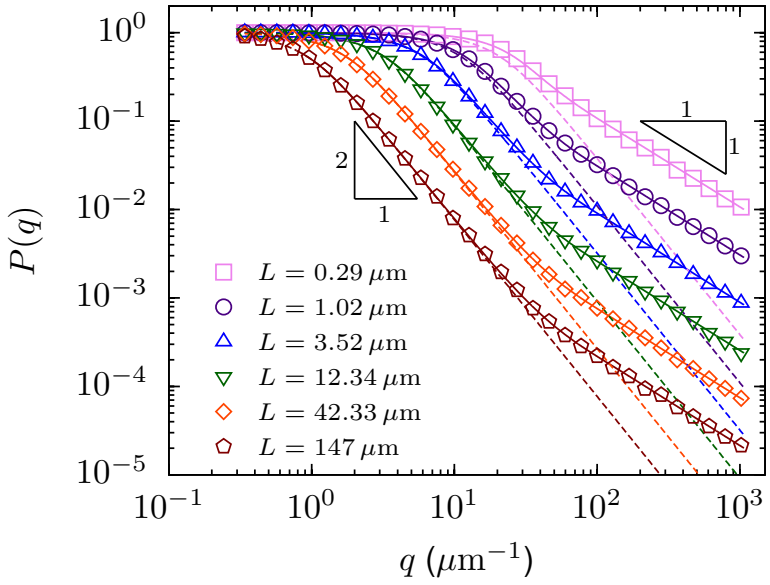

Figure A.3: Form factor for different lengths of dsDNA with no excluded volume. Solid lines correspond to ideal WLC expression in Eq. A.17, dashed lines show the curve obtained from the Debye expression for Gaussian chains and the data points correspond to the simulation results. Gaussian coil scaling $P(q) \sim q^{-2}$ for moderately high q and rod-like scaling $P(q) \sim q^{-1}$ for high q is clearly seen.

factor of wormlike chains with excluded volume, which appears as dotted lines in Fig. (5) of the main text. Using a modified form of the Daniel's distribution function [279] for stiff chains, the authors arrived at an expression for the form factor given by,

$$P(q) = I\left(\mathcal{N}\right) - I\left(k'\right) + \frac{2}{N^2}\int_0^k \mathrm{d}t(N-t)\left[1 - \frac{q^2\langle r^2\rangle}{3!} + \frac{q^4\langle r^4\rangle}{5!} - \ldots\right] \tag{A.19}$$

where N is the total number of beads, k is the number of beads where the modified Daniel's distribution gives the end-to-end distance of a wormlike chain with no excluded

volume [226] and $k' = kd/b$. Additionally, $I(y)$ is defined as

$$\begin{aligned} I(y) =& \frac{2}{\alpha^{2/(1+\epsilon)}\mathcal{N}(1+\epsilon)}\gamma\left(\frac{1}{1+\epsilon}, \alpha^2 y^{(1+\epsilon)}\right) - \frac{2}{\alpha^{4/(1+\epsilon)}\mathcal{N}^2(1+\epsilon)} \\ & \gamma\left(\frac{2}{1+\epsilon}, \alpha^2 y^{1+\epsilon}\right) + \frac{4(1-\exp(-\alpha^2 y^{1+\epsilon}))}{15\mathcal{N}(1+\epsilon)} - \frac{1}{\alpha^{2/(1+\epsilon)}\mathcal{N}^2(1+\epsilon)} \\ & \gamma\left(\frac{2+\epsilon}{1+\epsilon}, \alpha^2 y^{1+\epsilon}\right) + \frac{11\alpha^2 y^{1+\epsilon}\exp(-\alpha^2 y^{1+\epsilon})}{15\mathcal{N}(1+\epsilon)} + \frac{11}{15\alpha^{2/(1+\epsilon)}\mathcal{N}^2(1+\epsilon)} \\ & \gamma\left(\frac{3+2\epsilon}{1+\epsilon}, \alpha^2 y^{1+\epsilon}\right), \end{aligned} \tag{A.20}$$

where γ is the incomplete gamma function,

$$\gamma(a, x) = \int_0^x t^{a-1}\exp(-t)\mathrm{d}t \tag{A.21}$$

and α is a dimensionless form of the scattering vector

$$\alpha^2 = \frac{q^2 b^2}{6} \tag{A.22}$$

The parameter, ϵ, is defined by the ratio of the mean-square end-to-end distance of a polymer to its value in the absence of excluded volume, i.e., $\langle R^2\rangle_{\text{real}}/\langle R^2\rangle_{\text{ideal}} = N^\epsilon$.

The integral involving the moments of the pairwise distance $\langle r^n\rangle$ in Eq. A.19 was evaluated as follows:

$$\begin{aligned} & \frac{2}{N^2}\int_0^k \mathrm{d}t(N-t)\left[1 - \frac{q^2\langle r^2\rangle}{3!} + \frac{q^4\langle r^4\rangle}{5!} - \ldots\right] \\ & = \frac{2k'}{\mathcal{N}}\left[1 - \frac{k'}{2\mathcal{N}}\right] - \frac{q^2 b^2}{3\mathcal{N}^2}\left[k'^2\left(\frac{\mathcal{N}}{2} - \frac{k'}{3}\right) - \frac{k'}{2}\left(\mathcal{N} - \frac{k'}{2}\right)\right. \\ & - \frac{\exp(-2k')}{4}\left(\mathcal{N} - k' - \frac{1}{2}\right) + \frac{\mathcal{N}}{4} - \frac{1}{8}\Bigg] + \frac{q^4 b^4}{60\mathcal{N}^2}\left[\frac{5k'^3}{3}\left(\frac{\mathcal{N}}{3} - \frac{k'}{4}\right)\right. \\ & - \frac{26k'^2}{9}\left(\frac{\mathcal{N}}{2} - \frac{k'}{3}\right) - \frac{\exp(-6k')}{324}\left(\mathcal{N} - k' - \frac{1}{6}\right) + \frac{107}{54}\left(\mathcal{N} - \frac{k'}{2}\right) \\ & \left. + \exp(-2k')\left(\frac{5\mathcal{N}}{4} + \frac{\mathcal{N}k'}{2} - \frac{3k'}{2} - \frac{k'^2}{2} - \frac{3}{4}\right) - \frac{101\mathcal{N}}{81} + \frac{1457}{1944}\right] + \ldots \end{aligned} \tag{A.23}$$

As done in the original paper, we set $\epsilon = 0.11$. The value of k' at $\epsilon = 0.11$ was found by linear interpolation between the values at $\epsilon = 0.10$ and $\epsilon = 0.12$ taken from Table I in Appendix A of Sharp and Bloomfield [226]. Note that the assumption that the width

of DNA w is equal to the hydrodynamic diameter d is implicit in the aforementioned equations. In spite of this assumption we obtain good agreement between the form factor predicted from Eq. A.19 and from our simulations.

A.5 Diffusion of an Ideal WLC

Due to some typos in the original description of an the friction of an ideal WLC [129], we repeat the corrected equations here. In these equations note that $\sigma = 2.278$ and we have defined δ as the ratio d/b. For $L/b \leq \sigma$, the inner diffusivity is given by

$$\begin{aligned} D_{\text{inner}} = & \frac{k_B T}{3\pi\eta L}\left\{ C_1 \ln\left(\frac{L}{d}\right) + C_2 + C_3\left(\frac{L}{b}\right) + C_4\left(\frac{L}{b}\right)^2 \right. \\ & + C_5\left(\frac{L}{b}\right)^3 + C_6\left(\frac{L}{d}\right)^{-1}\ln\left(\frac{L}{d}\right) + C_7\left(\frac{L}{d}\right)^{-1} \\ & \left. + C_8\left(\frac{L}{d}\right)^{-2} + C_9\left(\frac{L}{d}\right)^{-3} + C_{10}\left(\frac{L}{d}\right)^{-4} + O\left[\left(\frac{L}{d}\right)^{-5}\right]\right\} \end{aligned} \tag{A.24}$$

and

$$C_1 = 1 - 0.01412\delta^2 + 0.00592\delta^4 + O(\delta^6) \tag{A.25}$$

$$C_2 = 0.3863 - 0.1667\delta + 0.0016\delta^2 - 0.0224\delta^3 - 0.0007\delta^4 + O(\delta^5) \tag{A.26}$$

$$C_3 = 0.1667 + 0.0222\delta^2 + 0.0017\delta^4 + O(\delta^6) \tag{A.27}$$

$$C_4 = 0.01883 - 0.00789\delta^2 - 0.00038\delta^4 + O(\delta^6) \tag{A.28}$$

$$C_5 = -0.002039 + 0.000805\delta^2 + 0.000017\delta^4 + O(\delta^6) \tag{A.29}$$

$$C_6 = 0.04167\delta + 0.00567\delta^3 + O(\delta^5) \tag{A.30}$$

$$C_7 = 0.5 + 0.0786\delta - 0.0094\delta^2 + 0.0107\delta^3 + 0.0039\delta^4 + O(\delta^5) \tag{A.31}$$

$$C_8 = -0.06250 + 0.00132\delta^2 - 0.00055\delta^4 + O(\delta^6) \tag{A.32}$$

$$C_9 = 0.001302\delta + 0.000181\delta^3 + O(\delta^5) \tag{A.33}$$

$$C_{10} = 0.001953 - 0.000064\delta^2 + 0.000027\delta^4 + O(\delta^6) \tag{A.34}$$

$$\tag{A.35}$$

For $L/b > \sigma$, the outer diffusivity is given by

$$D_{\text{outer}} = \frac{k_B T}{3\pi\eta L}\left[A_1\left(\frac{L}{b}\right)^{1/2} + A_2 + A_3\left(\frac{L}{b}\right)^{-1/2} + A_4\left(\frac{L}{b}\right)^{-1} + A_5\left(\frac{L}{b}\right)^{-3/2}\right] \tag{A.36}$$

with

$$A_1 = \frac{4}{3}\left(\frac{6}{\pi}\right)^{1/2} \tag{A.37}$$

$$\begin{aligned} A_2 = &-\left[1 - 0.01412\delta^2 + 0.00592\delta^4 + O(\delta^6)\right]\ln\delta \\ &- 1.0561 - 0.1667\delta - 0.1900\delta^2 - 0.0224\delta^3 + 0.0190\delta^4 + O(\delta^5) \end{aligned} \tag{A.38}$$

$$A_3 = 0.1382 + 0.6910\delta^2 \tag{A.39}$$

$$\begin{aligned} A_4 = &-\left[0.04167\delta^2 + 0.00567\delta^4 + O(\delta^6)\right]\ln\delta \\ &- 0.3301 + 0.5\delta - 0.5854\delta^2 - 0.0094\delta^3 - 0.0421\delta^4 + O(\delta^5) \end{aligned} \tag{A.40}$$

$$A_5 = -0.0300 + 0.1209\delta^2 + 0.0259\delta^4 \tag{A.41}$$

A.6 Comment on Single Molecule Experimental Data

In the main text we stated, but did not support, the idea that a constant change in length scale was not sufficient to explain the effect of fluorescent dyes on the diffusivity of dsDNA. In the case where the ratios b/w and b/d stay the same, a single simulation is sufficient to compare the change in the absolute length scale of the problem. In Fig. A.4(A), we compare the effect of the variation in the overall length scale on the PERM data (with excluded volume) on the diffusivity. As the figure depicts, the original scale describes the DLS and sedimentation data very well. An increase of 28% (expansion of the contour length of λ-DNA from 16.3 μm to 21 μm), increases the hydrodynamic radius, but not enough to fully describe the observed increase. To obtain a good match, the simulation requires an expansion near 100%, which corresponds to a contour length of ≈ 32 μm for λ-DNA.

In the main text, we also claim that the perceived slope of the diffusion coefficient is an insensitive indicator of the draining behavior of dsDNA. We support this by Fig. A.4(B), which shows the chain hydrodynamic radius as a function of contour length for over four orders of magnitude from both experimental data and PERM calculations.

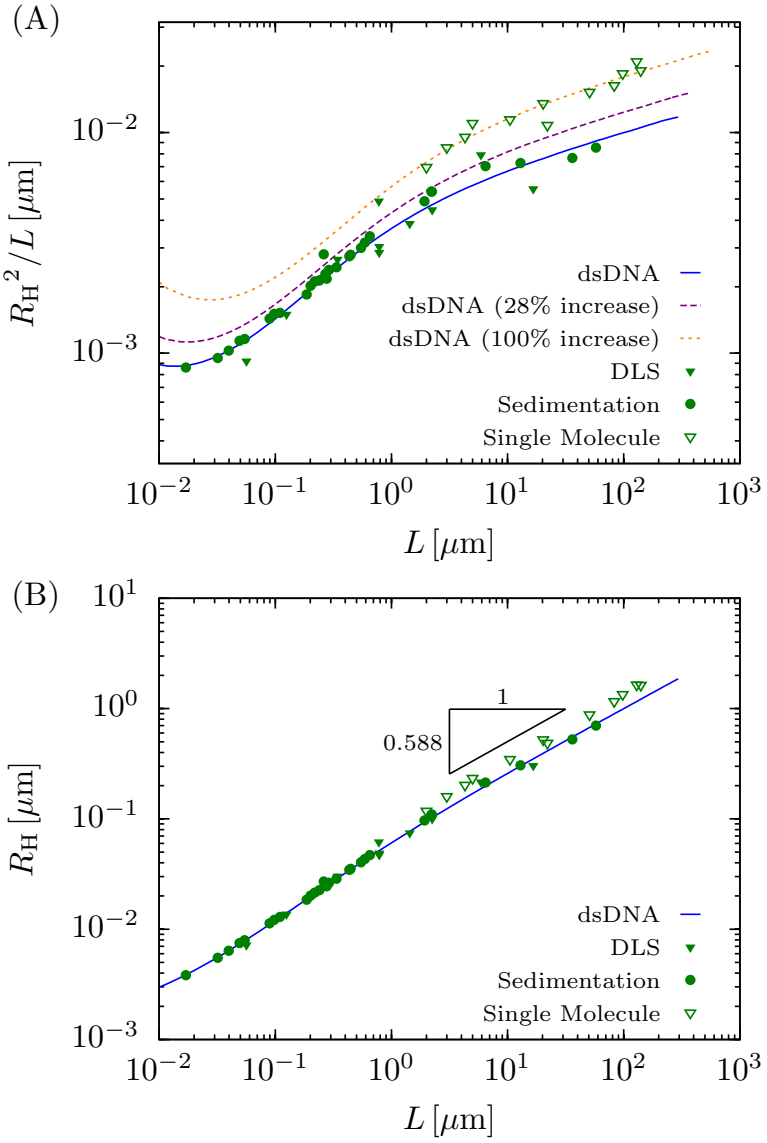

Figure A.4: (A) The square hydrodynamic radius as a function of contour length. The plot is normalized by the contour length to emphasize the variation in the data. (B) The hydrodynamic radius as a function of contour length. In both figures, experimental data appears as symbols and PERM simulations appear as lines. We have plotted the data in terms of the chain hydrodynamic radius, which allows us to compare experimental and simulation data regardless of solvent viscosity or ambient temperatures [89].

For reference we give the slope of 0.588, which in fact gives a moderately good description of most of the data on the plot, making it a poor measure of the draining behavior of the chain. While there is indeed some variation of the slope evident in the figure, it is quite subtle and difficult to detect with a least squares fit to noisy data over a limited range.

Appendix B

Supporting Information to Chapter 5

B.1 Simulation Method

In our implementation we have chosen an *off-lattice*, touching-bead model [75] in which we implement a discrete wormlike chain potential

$$\beta U_{\text{bend}} = \kappa \sum_{j=1}^{N_b-2} (1 - \mathbf{u}_{j+1} \cdot \mathbf{u}_j) \tag{B.1}$$

where $\beta = (k_B T)^{-1}$ is the inverse absolute temperature and $\mathbf{u}$ is the bond vector connecting bead j to bead $j+1$. The parameters in this equation are the number of beads (N_b), the bending potential constant (κ), and the bead size (w). In this model, choosing w is especially consequential because it (1) fixes the bond length, (2) sets the excluded volume to the value: $v \cong {l_p}^2 w$ (which is athermal because of hard beads and walls) and (3) imposes the length scale below which self-interactions are disallowed.

Simulations are run in either free solution or in an infinintely long square channel of size D. When confined, the channel walls are hard and are defined by the potential

$$\beta U_{\text{wall}} = \begin{cases} \infty & \max(|r_x|, |r_y|) \geq (D-w)/2 \\ 0 & \text{otherwise} \end{cases} \tag{B.2}$$

Here $|r_x|$ and $|r_y|$ are the absolute values of the chain position in the x and y directions respectively, D is the channel width and the z-axis is left open.

The bending potential is responsible for setting the value of the persistence length, which is defined as

$$\langle \mathbf{u}_{j+1} \cdot \mathbf{u}_j \rangle = \exp(-jw/l_p) \tag{B.3}$$

Fortuitously, the probability density function for a given bond angle defined as $\mathbf{u}_{j+1} \cdot \mathbf{u}_j = \cos\theta_j$ can be solved analytically [198] and is given by

$$P(\theta_j) = \frac{\kappa \exp\left[-\kappa(1-\cos\theta_j)\right]}{1-\exp(-2\kappa)}. \tag{B.4}$$

For illustrative purposes Fig. B.1 shows several plots of Eq. (B.4) for various values of κ.

The closed form solution in Eq. (B.4) is beneficial for two reasons. First, it allowed us to code PERM more efficiently, as the simulation requires generating many bond vectors according to this distribution. Second, it allowed us to obtain the persistence length for given values of κ and w. This can be done by calculating Flory's characteristic ratio [72, 199],

$$C_\infty = \frac{b}{w} = \frac{1+\langle\cos\theta_j\rangle}{1-\langle\cos\theta_j\rangle} \tag{B.5}$$

which gives the Kuhn length $b = 2l_p$. Evaluating this expression leads to [198]

$$l_p = \left(\frac{w}{2}\right)\frac{\kappa - 1 + \kappa\coth\kappa}{\kappa + 1 - \kappa\coth\kappa} \tag{B.6}$$

which simplifies to the well-known expression for the persistence length,

$$\frac{l_p}{w} = \kappa - 1/2 \tag{B.7}$$

when $\kappa \gg 1$.

In addition to using an analytical distribution to generate the bond vectors, we further accelerated the simulations by taking advantage of the fact that the confinement free energy is extensive in N for sufficiently long chains. PERM relies on an estimate of the partition function in order to bias the sampling. In the "blind" version of the PERM algorithm [173], the estimate is progressively built at the beginning of each simulation run, a process which dominates the simulation time. To speed up the convergence of the estimate, we first ran a blind simulation for short chains and linearly extrapolated the partition estimate to get a good initial guess for the partition function for the longer chains. This guess is only used to set the biasing in PERM, and does not affect the final

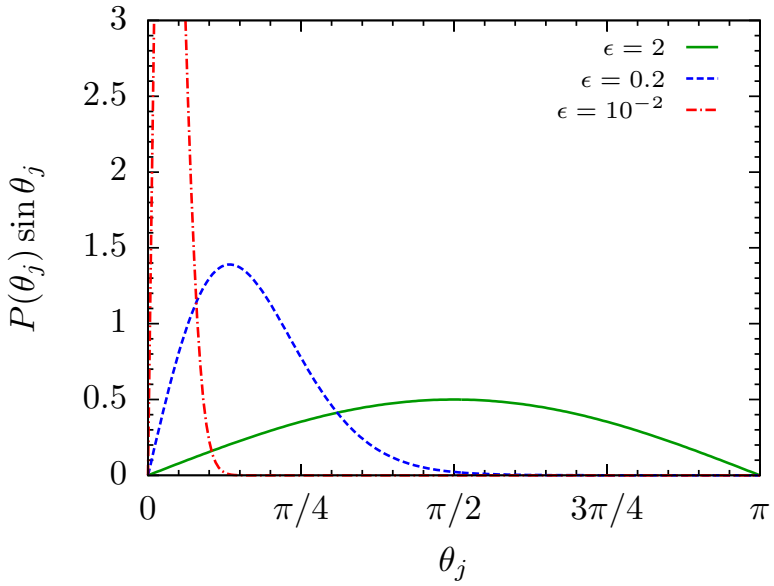

Figure B.1: Various curves of Eq. (B.4) for three values of κ: 10^{-5} (solid), 5.499 (dashed), 100.5 (dash-dot). These values illustrate the bond angle probability density function when chains are very stiff ($\kappa = 100.5$) or very flexible ($\kappa = 10^{-5}$, practically a freely-joined chain).

partition function estimate. We were able to run "non-blind" simulations [173] for the longest chains, a technique which reduced the simulation time considerably.

In our implementation, we employed a master/slave parallel algorithm *without* Markovian anticipation [200] on a DELL linux cluster. (Markovian anticipation is not trivial to implement with an off-lattice model.) The hardcore excluded volume interaction calculations took advantage of neighbor lists, and data analysis was done on the fly since recording each tour's configuration is prohibitively expensive. Most of the PERM data (unless indicated otherwise) was taken for chains with 2×10^4 beads in five batches of 10^4 tours for error estimation; in all cases the error of the data shown is smaller than the symbol size. Our simulations spanned the wide range of channel sizes and chain anisotropies shown in Fig. B.2, corresponding to four decades in the dimensionless channel size $\delta \equiv D/l_p$ and three decades in the anisotropy $\epsilon \equiv w/l_p$. In contrast, most previous work, focusing on DNA in nanochannels [71, 139, 140, 142, 237, 238], spans less than two decades in channel size (say, 10 nm to 500 nm) and uses a single value of the anistropy ϵ corresponding to DNA or, at most, a change in ϵ by a factor of around 5. Moreover, these previous simulations typically use around 10^3 beads to represent the DNA.

The free energy of a given chain was obtained from PERM's estimate of the canonical partition function [171] which has an unconfined, ideal wormlike chain standard state ($\epsilon = 0$). Thus, the confinement free energy

$$\Delta F_c \equiv F_{\text{confined}} - F_{\text{bulk}} \tag{B.8}$$

requires simulations of both confined and free solution chains for chains with non-zero excluded volume. In order to assure that the free energy calculations were accurate, we verified that all of the free energy calculations shown in the main text did in fact reach the asymptotic limit where $\mathcal{F} \sim N^0$ where $N \equiv L/l_p$ is the number of persistence lengths of the chain and $\mathcal{F} \equiv \beta \Delta F_c N^{-1}$ is the dimensionless free energy. Figure B.3 shows an example set of data ($\delta_{\text{eff}} = 9.9$ and $\epsilon = 0.1$) which corresponds to DNA in a high ionic strength buffer in a 500 nm channel. As we can see, the simulation spans 1000 persistence lengths, corresponding to a contour length of 50 μm, or about 150 kbp (slightly smaller than T4 DNA). The plateau for $\mathcal{F}$ in Fig. B.3 is the long-chain asymptotic value for this particular combination of δ_{eff} and ϵ appearing in the main

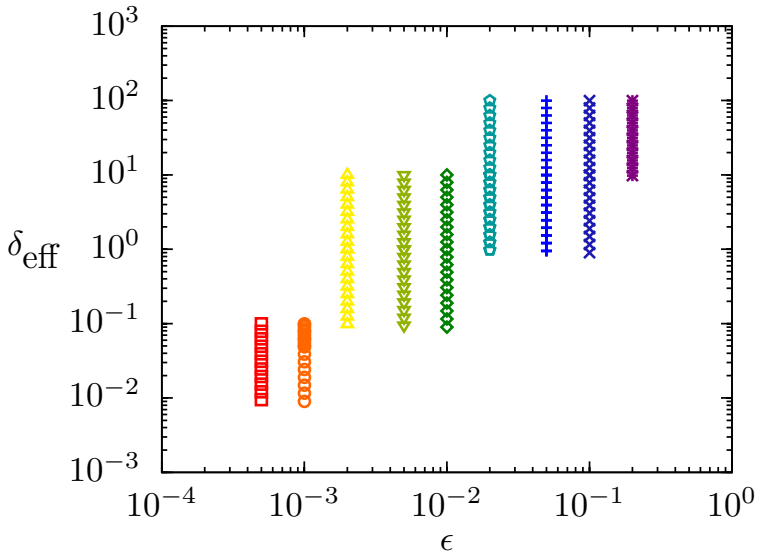

Figure B.2: Range of values for the effective channel size available to the chain, $\delta_{\text{eff}} \equiv (D - w)/l_p$, and the chain anisotropy, $\epsilon \equiv w/l_p$, used in the PERM simulations. The phase space explored here is orders of magnitude larger than previous studies. The umbrella-like overlap between different values of ϵ allows us to produce the universal free energy curve in Fig. 4 of the main text.

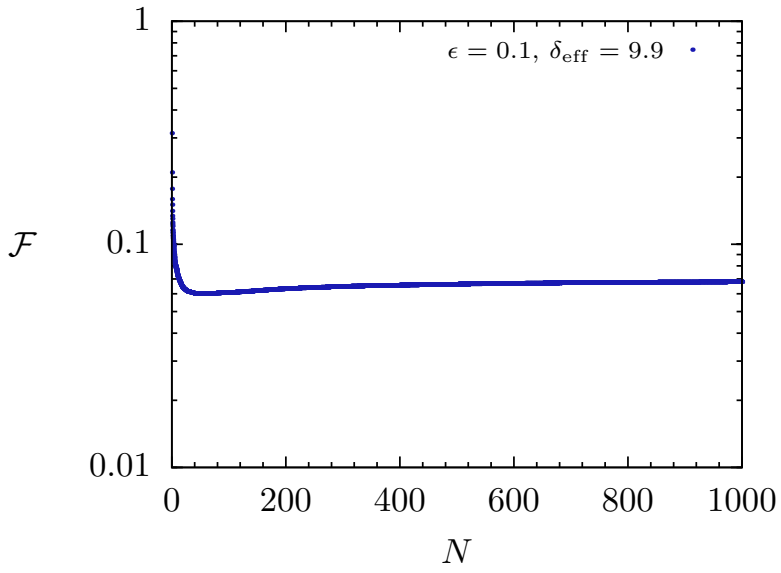

Figure B.3: Example of the convergence of the confinement free energy, $\mathcal{F} \equiv \beta \Delta F_c N^{-1}$ as a function of the size of the chain, $N \equiv L/l_p$. The data appearing in Fig. 4 of the main text correspond to the plateau region for a given simulation for a channel size δ and chain anisotropy ϵ.

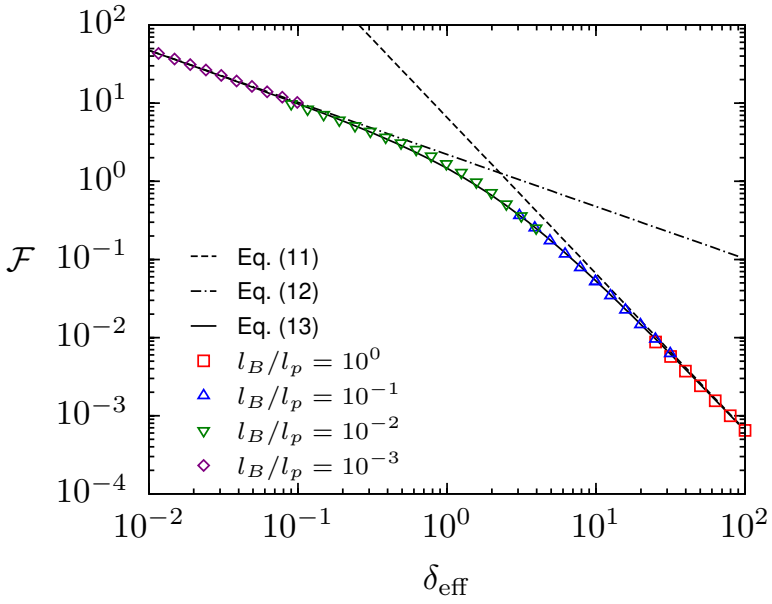

Figure B.4: Free energy of a confined semiflexible chain without excluded volume. Simulation data is taken for chains of different persistence length where the bond length is represented as l_B. For each simulated chain $\epsilon = 0$, except for $l_B/l_p = 10^{-3}$, which has a small but negligible $\epsilon = 10^{-3}$. The data are used to validate Eq. (B.9) and calculate C_1 by a least-squares regression.

text. We constructed similar plots for every combination of δ_{eff} and ϵ and included only those simulations which reached a plateau value.

B.2 Confinement Free Energy of an Ideal Semiflexible Chain

Following Chen and Sullivan [240] we propose an interpolation formula of the form

$$\mathcal{F} = \frac{\frac{2}{3}\pi^2\delta^{-2}}{(C_2\delta^{-2} + C_1\delta^{-1} + 1)^{2/3}} \tag{B.9}$$

The choice $C_2 = 5.147$ matches previous calculations [132] for the Odijk regime. The remaining constant is used to fit the shape of the crossover region from PERM simulations in the absence of any excluded volume. A least squares fit yields $C_1 = 3.343$ and the interpolation shown in Fig. B.4.

B.3 Numerical Prefactors for the Odijk and Flory-de Gennes Extension

In several instances in the main text, PERM results for the extensions are compared to Odijk and de Gennes theories with exact prefactors. The Odijk regime curve corresponds to

$$\langle X/L \rangle = 1 - 0.18274 \left(\frac{D-w}{l_p} \right)^{2/3} \tag{B.10}$$

which is the prediction of the Odijk theory [58] using the prefactor computed by Burkhardt et al. [132] for a square nanochannel. The quantity $D - w$ is the effective width of the nanochannel available to the chain.

For the Flory-de Gennes regime, the extension is

$$\langle X/L \rangle = (1.033 \pm 0.005) \left(\frac{D-w}{l_p} \right)^{1-1/\nu} \left(\frac{w}{l_p} \right)^{2-1/\nu} \tag{B.11}$$

where the prefactor is obtained by a fit to the collapsed PERM data shown in Fig. B.5. The prefactor of almost unity is a satisfying test of the de Gennes theory for a finite width chain, derived above.

In Fig. 3 of the main text, the data shown corresponds to $\epsilon = 10^{-1}$ and $\epsilon = 10^{-2}$ for a range of values of $\delta \equiv D/l_p$. These dimensionless data were converted using a persistence length $l_p = 50$ nm and two different values of the width, $w = 5$ nm and $w = 0.5$ nm. The corresponding channel sizes D follow from the definition of δ.

We also stated that the Odijk regime is valid for an effective channel size $\delta_{\text{eff}} \equiv (D - w)/l_p \leq 0.3$ and the Flory-de Gennes regime begins at $\delta_{\text{eff}} \geq (2\epsilon)^{-1}$. The approximate numerical values 0.3 and 1/2 were obtained by inspection of the data in Fig. B.5A and B.6A in the region where the data appear to collapse onto the universal curves. Additionally, one can again note that these regimes show the lack of collapse to the Odijk theory for a fractional extension below 90%. A similar general statement is not possible for the de Gennes theory, since this value depends on ϵ, as is evident in Fig. B.6A.

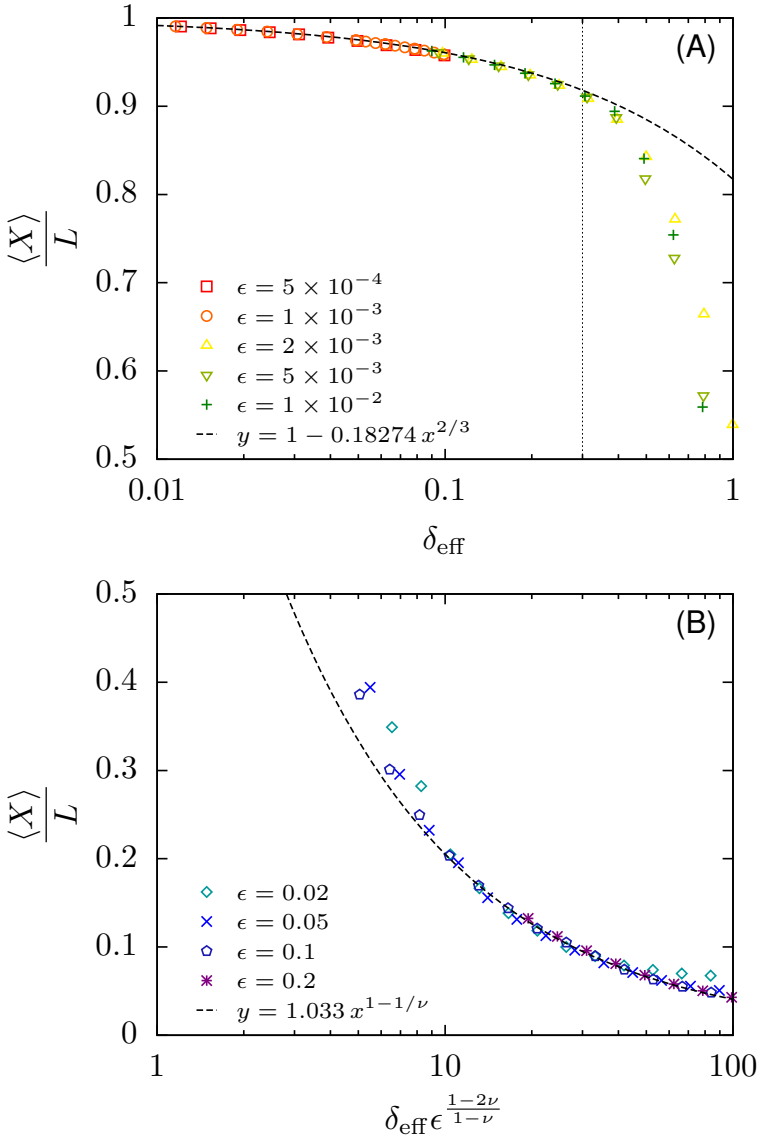

Figure B.5: Fractional extension of a semiflexible chain confined in a square nanochannel collapsed to the (A) Odijk regime and (B) Flory-de Gennes regime. (A) The data collapses before $\delta_{\text{eff}} = 0.3$ as indicated by the dotted vertical line. (B) Note here that the dependent variable is set to $x = \delta_{\text{eff}} \epsilon^{\frac{1-2\nu}{1-\nu}}$ and that $x^{1-1/\nu}$ yields the right hand side of Eq. (B.11). This was done to make the ordinate linear in the channel size.

B.4 Scaling for the extension in the quasi-Gaussian regime

Figure 3 of the main text shows the extension data on a linear plot in dimensional units, which is the easiest way to make a connection to the experimental data. Figure B.6B shows the same data in a log-log plot using the effective dimensionless channel size, δ_{eff}, for the abscissa. These data correspond to asymptotically long chains; for 2×10^4 beads of size 5 nm, these data are for chains 100 μm long (approximately 300 kilobase pairs). Our extension results complement previous work [71, 139, 140, 142, 238] that used chains that are several microns long.

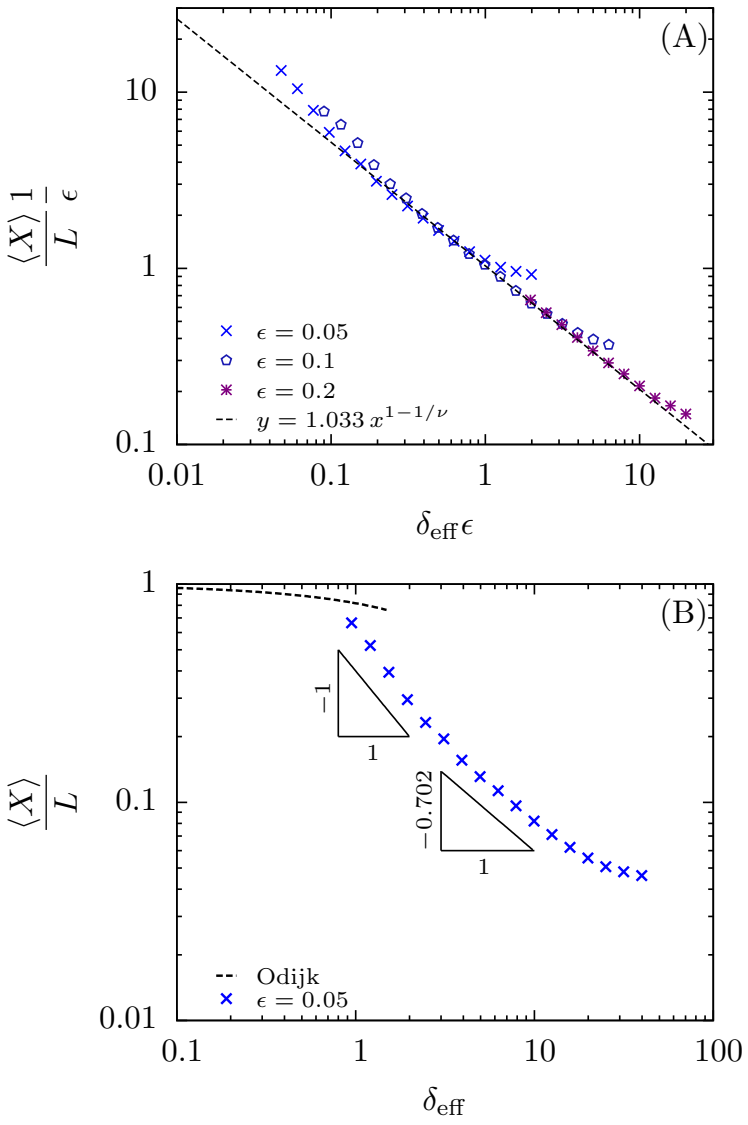

Figure B.6: (A) Log-log plot of the fractional extension from Fig. B.5B, which shows the collapse to the Flory-de Gennes regime around $\delta_{\rm eff}\epsilon \approx 1/2$. This corresponds to about 200 nm for DNA ($\epsilon = 0.1$). (B) Average fractional extension versus dimensionless channel size for DNA ($\epsilon = 0.1$) obtained from PERM simulations. Both the Flory-de Gennes scaling and the Gauss-de Gennes scaling are indicated.